Spatial protein interaction networks of the intrinsically disordered transcription factor CEBPA

Dissertation
zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

im Fach Biologie/Molekularbiologie

eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Von

Evelyn Ramberger, M.Sc.

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing.Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachter:

1. Prof. Dr. Achim Leutz
2. Prof. Dr. Matthias Selbach
3. Prof. Dr. Gunnar Dittmar

Tag der mündlichen Prüfung: 12.8.2020

For T.

Table of Contents

Selbstständigkeitserklärung

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben.
Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad.
Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde.
Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin vom 5.März 2015.
Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsbearbeiter*innen stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, 4.3.2020

Evelyn Ramberger

List of Figures

List of Tables

Abbreviations

AA	amino acid
ABC	ammonium bicarbonate
Ac	acetylation
ACN	acetonitrile
AML	acute myeloid leukaemia
APL	acute promyeloid leukaemia
AP-MS	affinity purification coupled to mass spectrometry
ATRA	all-trans retinoic acid
BirA	biotin ligase
BioID	proximity dependent biotin identification
bZIP	basic leucine zipper
C/EBP	CCAAT/enhancer binding protein
Citr	arginine citrullination
CR	conserved region
DB	DNA binding region
DNA	deoxyribonucleic acid
Dox	doxycycline
FA	formic acid
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FDR	false discovery rate
GSEA	gene set enrichment analysis
GO	gene ontology
HAT	histone acetyltransferase
HDAC	histone deacetylase
H/M/L	heavy/medium/light SILAC labels
HPLC	high performance liquid chromatography
iBAQ	intensity based absolute quantification
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IT	injection time
LC	liquid chromatography
LC MS/MS	liquid chromatography coupled to tandem mass spectrometry
LFQ	label free quantification
LysC	lysyl endopeptidase
LZ	leucine zipper
Me	methylation
Me2	dimethylation
Me2 sym/asym	symmetric and asymmetric arginine dimethylation
min	minute
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS1	peptide mass-to-charge ratio scan
MS2	peptide fragmentation spectrum
msec	millisecond
m/z	mass-to-charge ratio

PBS	phosphate buffered saline solution
PCA	principal component analysis
PML	promyelocytic leukaemia
PRISMA	protein interaction screen on peptide matrix
PRM	parallel reaction monitoring
PRMT	protein arginine methyltransferase
PTM	post-translational modification
RD	regulatory domain
RT	room temperature
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
SILAC	stable isotope labelling with amino acids in cell culture
TAD	transactivation domain
TE	transactivating element
TFA	trifluoracetic acid
TPA	12-O-Tetradecanoylphorbol-13-acetate
TRE	tetracycline response element
WT	wild type

Zusammenfassung

Der Transkriptionsfaktor C/EBPα reguliert Differenzierung und Proliferation in verschiedenen Zelltypen und spielt eine herausragende Rolle in der Hämatopoese. Die CEBPA RNA kann in die lange P42-Isoform oder die N-terminal verkürzte P30-Isoform translatiert werden. Während P42-C/EBPα differenzierungsinduzierend wirkt, ist P30 als Inhibitor von P42 und als Onkogen in akuter myeloider Leukämie beschrieben. Die Modularität und Multifunktionalität von C/EBPα, die ihn zahlreichen Studien beobachtet wurde, lässt sich möglicherweise durch differentielle Protein–Protein-Interaktionen erklären. Zahlreiche post-translationale Modifikationen (PTMs) und die intrinsisch ungeordnete, flexible Struktur von C/EBPα stellen jedoch eine Herausforderung für traditionelle Ansätze in Proteininteraktionsstudien dar. In der vorliegenden Arbeit wird ein neuer, alternativer Ansatz präsentiert, der auf einem *in vitro* Proteininteraktions-screen auf einer Peptidmatrix (PRISMA) und Biotinligase proximity labelling (BioID) in lebenden Zellen basiert. In einem PRISMA-screen wurden 120 C/EBPα Peptide auf Proteininteraktionen mit Proteinextrakt aus myeloiden Zellen untersucht. Im Screen wurden 40 verschiedene C/EBPα PTMs inkludiert, unter anderem auch die hier erstmals neu beschriebenen Methylierungen der C/EBPα Argininreste R12 und R142. Daten aus dem PRISMA-screen wurden mit BioID Experimenten in myeloiden Zellen validiert, um eine Proteininteraktionslandkarte von C/EBPα zu generieren, die 52 bekannte und 68 neue C/EBPα Proteininteraktoren umfasst. Hotspots für Proteininteraktionen fallen in evolutionär konservierte C/EBPα Regionen und der Vergleich des Bindungsprofils mit publizierten Daten zeigt Ähnlichkeiten zu verwandten Transkriptionsfaktoren der C/EBP Familie. Die Daten aus PRISMA und BioID deuten darüber hinaus an, dass die P30-C/EBPα Isoform kein bloßer Inhibitor von P42-C/EBPα ist, sondern eine transkriptionell schwächere Variante von C/EBPα mit modulierter Funktion darstellt. Die Ergebnisse legen nahe, dass die Multifunktionalität von C/EBPα von multivalenten Proteininteraktionen in Abhängigkeit von PTMs koordiniert wird um C/EBPα mit dem epigenetischen und transkriptionellen Apparat der Zelle verknüpfen.

Summary

The pioneering transcription factor C/EBPα plays a lineage-instructing role during haematopoiesis and also regulates proliferation and differentiation in many other cell types. The CEBPA RNA can be translated into a full length (P42-C/EBPα) or N-terminally truncated isoform (P30-C/EBPα). While P42 induces differentiation in various cell types, the P30 isoform is mostly regarded as a dominant inhibitor of P42-C/EBPα and acts as an oncogene in acute myeloid leukaemia. Protein interactions may be the key to explaining the functional plasticity and modularity of C/EBPα that has been demonstrated in diverse experimental settings. However, the disordered structure and the numerous post-translational modification sites (PTMs) of C/EBPα pose a challenge to traditional protein interaction studies. In the present work, a novel alternative approach is presented that combines an in vitro protein interaction screen on a peptide matrix (PRISMA) with biotin ligase proximity labelling (BioID) in living cells. To this end, 120 C/EBPα peptides were probed for protein interactions with PRISMA. The screen comprised 40 different PTMs, including the newly identified C/EBPα arginine methylation sites R12 and R142. PRISMA data was validated with BioID experiments and generated a detailed C/EBPα protein interaction map in myeloid cells. The interactome presented here contains 52 known and 68 novel C/EBPα interactors that can now be mapped across the C/EBPα sequence in a PTM dependent fashion. Hotspots of protein interaction correlated with conserved regions and comparison with previously published data revealed related binding profiles of homologous C/EBP regions. The interactome data furthermore hints that P30 does not act as a dominant suppressor of P42 but rather as a transcriptionally weaker derivative that contains unique properties and lacks some features of P42. Taken together, the data indicates that the functional plasticity of C/EBPs is orchestrated by multivalent protein interaction events and PTMs to configure a dynamic C/EBP hub that interacts with many partners of the transcriptional and epigenetic machinery. The experimental strategy of combining PRISMA with BioID may serve as a basis to explore the linear and PTM-dependent interactome of a vast number of intrinsically disordered proteins involved in cell signalling and gene regulation.

1. Introduction

1.1. Disordered proteins

For many years the understanding of proteins was shaped by the dogma that a fixed three-dimensional protein structure, determined by the primary amino acid sequence, mediates protein functions. In a broader sense this is exemplified by the textbook “lock and key” model introduced by Emil Fischer over a hundred years ago that states that enzymes bind to their substrates because their shapes fit perfectly (Fischer, 1894). Appropriately, the alteration of protein structure during denaturation leads to the loss of enzymatic activity. Since the publication of the first three-dimensional protein structure of myoglobin around 60 years ago (Kendrew et al., 1958), there has been a vast increase in our knowledge of protein structures and with it a notion that protein folding is inherently connected to function. While this functional link is undeniable, in the last three decades it became clear that a significant fraction of the eukaryotic proteome is either fully disordered (intrinsically disordered proteins, IDPs) or contains stretches of disordered regions (intrinsically disordered regions, IDRs). In this thesis I will use IDP as a generic term to denote proteins containing extensive disorder.

Disordered protein sequences are unable to spontaneously fold into a stable three-dimensional structure and are characterized by conformational flexibility that enables them to rapidly alternate between a wide range of possible conformations (Dyson and Wright, 2005). Since the mid-1990s, nuclear magnetic resonance spectroscopy (NMR), as well as circular dichroism, fluorescence spectroscopy and X-ray experiments have provided indisputable *in vitro* evidence for IDPs (Daughdrill et al., 1997; Dyson and Wright, 2004; Kriwacki et al., 1996; Li and Song, 2007; Wells et al., 2008; Wright and Dyson, 1999; Zhang et al., 1994). NMR and single molecule fluorescence resonance energy transfer (FRET) measurements in living cells have suggested that some IDPs also remain unstructured in the crowded environment of the cell (Leblanc et al., 2018; Luchinat and Banci, 2016).

The amino acid composition of IDPs is of low sequence complexity and is biased towards a high content of polar and charged amino acids as well as a low fraction of bulky, hydrophobic amino acids that prevents the formation of a hydrophobic core (Garner et al., 1998; Ferron et al., 2006). Several bioinformatic tools have been developed to predict disorder probability of a given amino acid sequence, either based on sequence composition or machine learning approaches trained on experimental data (He et al., 2009; Nielsen and Mulder, 2019). According to current estimations, around 15% of all human proteins are fully disordered and another 35% contain

disordered regions of at least 30 amino acids (Forman-kay and Mittag, 2013). While some proteins are completely disordered, many proteins like C/EBP transcription factors contain disordered regions together with more structured domains **(Figure 1)**. Compared to prokaryotes, the eukaryotic proteome has a significantly higher fraction of IDPs, giving rise to speculations that disorder is connected to evolutionary complexity (Dunker et al., 2000; Ward et al., 2004).

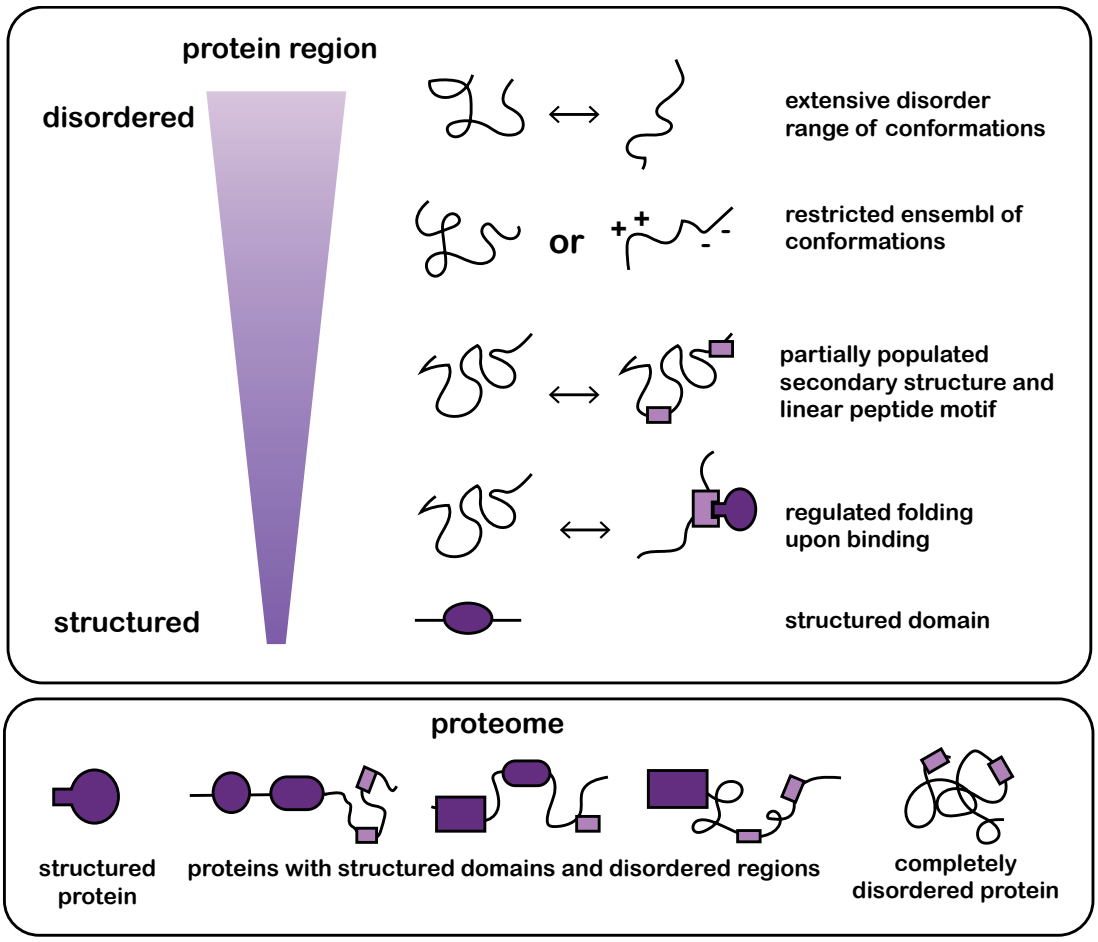


Figure 1: Disordered and structured protein regions.
Intrinsically disordered regions are characterised by a flexible structure that allows a range of conformations. Many proteins contain structured together with disordered regions. This figure was adapted from Babu et al., 2012.

1.1.2. Functions of disordered proteins

IDPs play a pivotal role in regulation of gene expression and cell signalling (Iakoucheva et al., 2002; Liu et al., 2006; Ward et al., 2004; Wells et al., 2008). Other IDP functions include but are not limited to: protein phosphorylation, nucleic acid and small molecule binding and self-assembly of multi-subunit protein complexes (Iakoucheva et al., 2004; Metallo, 2011; Simone et al., 2012; Varadi et al., 2015). Exciting discoveries from the last couple of years also indicate that disordered proteins can drive phase separation to form membrane-less organelles within the nucleus or cytoplasm (Boeynaems et al., 2018).

The structural flexibility of IDPs is instrumental to their functional plasticity and enables them to participate in promiscuous interactions with different target molecules. In the following paragraphs I will discuss protein-protein interactions mediated by disordered proteins and their implications in gene regulation in more detail. A comprehensive overview over the broad functions of IDPs is beyond the scope of this thesis and is given in the reviews of Forman-kay and Mittag, 2013 and Wright and Dyson, 2015.

1.1.3. Protein interactions mediated by disordered regions

IDPs often interact with multiple binding partners and are located in the center of protein interaction networks where they act as hub proteins (Haynes et al., 2006; Kim et al., 2008). Such hub proteins frequently contain unstructured disordered regions together with structured domains that are both involved in mediating protein interactions. A well-studied example of such a hub protein is the histone acetyltransferase EP300 that regulates gene expression through chromatin remodelling. Around 50% of the amino acid sequence of EP300 is disordered and it is estimated that it has up to 400 different interaction partners; some of them, like HIF1α, interact with structured parts of P300 while some interactions are mediated by intrinsically unstructured parts of EP300 (Dyson and Wright, 2016).

While some IDPs retain disorder also in complex with an interaction partner (“fuzzy complexes”), many IDPs undergo structural changes upon binding (Mollica et al., 2016). Such coupled binding and folding events have been described for a number of proteins, one of the most studied examples is the binding of the transcriptional regulator cAMP response-element binding protein (CREB) to its co-activator CREB binding protein (CBP) (Dyson and Wright, 2016). Protein regions that undergo disorder-to-order transition upon binding events are also called molecular recognition features (MoRFs) and are usually around 20 amino acids long. Shorter disordered binding

regions consisting of 3 – 12 residues are called short linear motifs (SLiMs). According to current estimations based on SLiM and MoRF prediction tools, the human proteome might contain as many as 132000 of these disordered protein interaction motifs (Tompa et al., 2014)

The relatively small binding interfaces in SLiM-mediated interactions result in low to moderate affinities with high dissociation rates, allowing for rapid and dynamic binding events (Babu, 2009; Christensen and Klevit, 2009; Ivarsson and Jemth, 2019). Nevertheless, SLiM based interactions can be highly specific, raising the conceptual question of how specificity in this context is achieved. Although a motif may be as short as three amino acids, the flanking regions and local context of the core binding motif can enhance specificity of the protein interaction (Stein and Aloy, 2008). SLiMs frequently contain sites for post-translational modification (PTMs) that alter functionality of motifs and around 13% of ligand binding motifs in the eukaryotic linear motif database (ELM) are regulated by PTMs (Davey et al., 2012). PTMs can either act like an “on/off switch” for an interaction like in the interaction of the CTiP complex with BRCA (Varma et al., 2005), or shift the specificity of a motif from one domain to another like the phosphorylation controlled internalization of CLTA-4 (Shiratori et al., 1997). In addition, the larger structural context of a motif may influence specificity and affinity. IDPs can contain several recognition motifs that interact with two or more domains of a single protein or with different subunits of a protein complex (Barbar and Nyarko, 2015; Clark et al., 2018). The high local concentration of another binding site upon dissociation from the initial site increases avidity in such multivalent interactions (Kitov and Bundle, 2003). Allovalency is an extension of the multivalency concept. Here, specificity and affinity are enhanced by several motifs that bind to the same target site in the interacting protein (**Figure 2**).

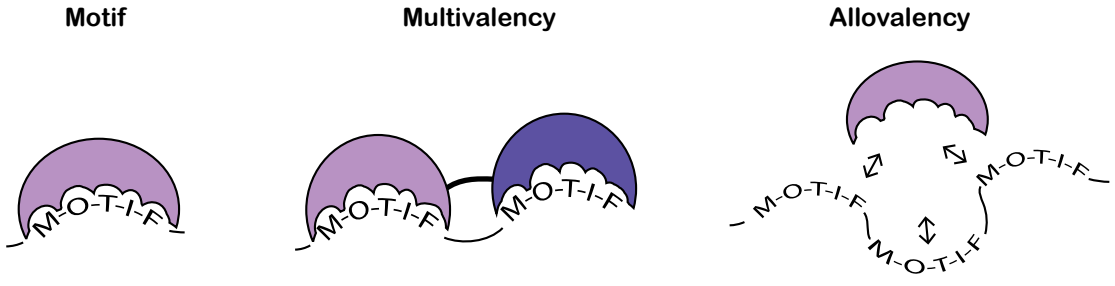


Figure 2: Motif based protein interaction.
Schematic representation of a protein interaction mediated by a disordered binding motif. Multivalency and allovalency increase affinity and specificity of the interaction by providing a secondary binding site with high local concentration. This figure was adapted from Ivarsson and Jemth, 2019.

Such multivalent interactions likely play a role in fine-tuning the interaction between sequence specific transcription factors and coactivators or repressors (Clark et al., 2018). An example for such a multivalent transcription factor is the ETS transcription factor ETV4. ETV4 contains several aromatic-rich motifs in the activation and DNA binding domain that interact with three different sites of the mediator of transcription complex (MED) subunit MED25 (Currie et al., 2017). A comparative analysis revealed that 82–94% of transcription factors contain extended regions of intrinsic disorder, predominantly located in transactivating regions (Liu et al., 2006). Furthermore, another study proposed that in transcription factors containing IDRs, alternative splicing and PTMs work together to provide a complex and context-specific toolkit for gene regulation (Zhou et al., 2018). Consequently, individual regions of transcription factors can confer different different functions that are also maintained outside of the structural context of the whole protein. This modular organisation of transcription factors was recognized as early as 1985 by domain swapping experiments with Lex4 and Gal4 (Brent and Ptashne, 1985) and since then has been confirmed in many different scenarios (Andreasson and Ljungdahl, 2004; Majello et al., 1997; Seipel et al., 1992; Xu et al., 2018).

1.2. C/EBP transcription factors

The family of CCAAT enhancer binding proteins contains six proteins (C/EBPα,β,γ,δ,ε,ζ) and represents an example for modular transcription factors that contain extensive IDRs. C/EBPs are involved in differentiation and cell fate decisions in various cell types, including myeloid cells, adipocytes and hepatocytes (Lekstrom-Himes and Xanthopoulos, 1998). In the following paragraphs I will give an overview about C/EBP structure and biology with a focus on C/EBPα.

1.2.1. Structure of C/EBPs

Structurally, C/EBP transcription factors represent modular proteins with extended intrinsically disordered regions in the N-terminus. All members of the C/EBP family contain a C-terminal leucine rich domain (leucine zipper, LZ) for homo or hetero dimerisation and an adjacent basic region (BR) for DNA binding (Lekstrom-Himes and Xanthopoulos, 1998; Ramji and Foka, 2002). The complete region of BR and LZ is addressed as basic leucine zipper domain (bZIP) and is also contained in several other transcription factor families. Dimerisation of bZIP transcription factors is a prerequisite for DNA binding and, through the amount of possible heterotypic combinations, increases the functional plasticity of these proteins (Amoutzias et al., 2007). The N-terminal part of C/EBP factors is more variable than the highly conserved bZIP domain. It contains several transactivating and regulatory regions that are conserved between species and share sequence homologies across different C/EBP transcription factors as indicated in **Figure 3B**.

In C/EBPα several conserved regions alternate with regions of low complexity (**Figure 3C**). In previous studies, N-terminal transactivating regions of C/EBPα are often referred to as transactivating domains (TADs) or transactivation elements (TEs). Alignment of C/EBP protein sequences from different vertebrate species revealed several conserved regions (CRs) within the TAD and TE regions (represented as colored boxes in Figure 3, adapted from Leutz et al., 2011). While the previously annotated TAD and TE regions span up to 80 residues, CRs are significantly smaller and vary between 7 and 28 amino acids (**Figure 3A**). In the present work, I will use the CR terminology to address individual C/EBPα regions.

Functional plasticity of C/EBPα, and C/EBP proteins in general, is increased though the production of different isoforms that regulate different aspects of C/EBP biology. C/EBPα,β and ε are expressed as different isoforms with distinct biological functions (Lekstrom-Himes and Xanthopoulos, 1998; Ramji and Foka, 2002). The intronless CEBPA RNA can be translated into two different isoforms via an alternative

translation initiation site. The N-terminally truncated C/EBPα isoform P30 lacks the first 119 amino acids present in the full-length P42 isoform (**Figure 3A**). Since P30-C/EBPα lacks major N-terminal transactivating regions, it is generally considered a dominant inhibitor of full length C/EBPα. However, an additional gene regulatory function towards the middle of the protein is also contained in the truncated P30 (TEIII or TADII).

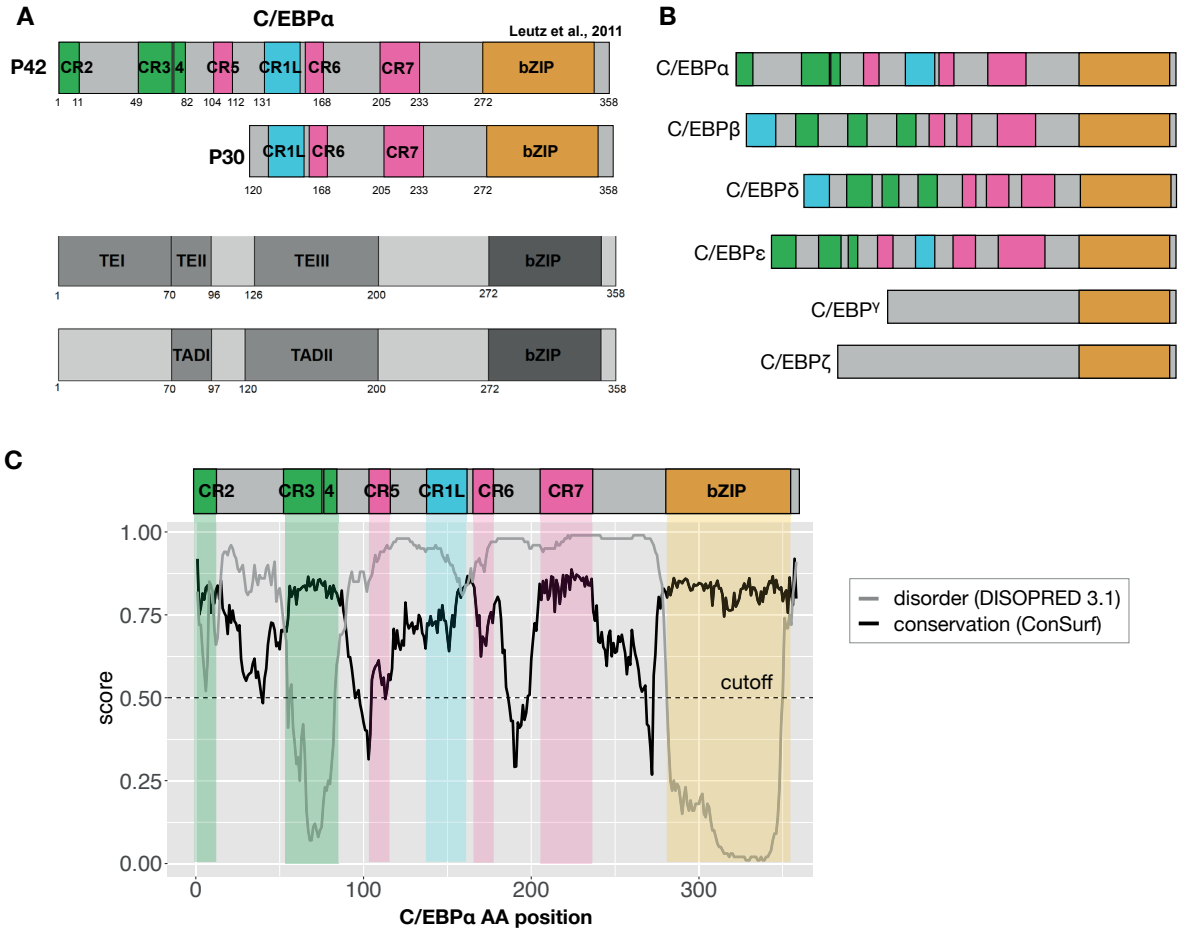


Figure 3: C/EBPα is an intrinsically disordered and modular protein.
A: Different annotations exist for C/EBPα: Conserved regions (CR, adapted from Leutz et al., 2011), transactivating elements (TEs) and transactivating domains (TADs) **B**: Members of the C/EBP transcription factor family share a basic leucine zipper domain (bZIP) and contain homologous conserved regions (CRs) in the N-terminus. **C**: Disorder prediction and sequence conservation across C/EBPα sequence. Disorder was calculated with DISOPRED v3.1, sequence conservation was calculated by comparing C/EBPα sequences from five vertebrate species (human, rat, chicken, cow, frog) with the ConSurf algorithm. Disorder threshold of 0.5 is indicated with a dotted line.

1.2.2. Biological role of C/EBPα

C/EBPα is the founding member of the C/EBP family and has first been isolated and cloned from rat liver more than three decades ago (Graves et al., 1986; Johnson et al., 1987). C/EBPα functions as a pioneering transcription factor that can directly bind to target sites in condensed chromatin and recruit other transcription factors and chromatin modifying enzymes to regulate expression of cell type specific genes (Madsen et al., 2014; Zaret and Carroll, 2011). Ectopic expression of C/EBPα induces differentiation of several cell types *in vitro* including macrophages, granulocytes and adipocytes (Porse et al., 2001; Radomska et al., 1998). Additionally, C/EBPα expression is capable of transdifferentiating B-cells into myeloid cells (Huafeng et al., 2004) and enhances reprogramming efficiency of B-cells into induced pluripotent stem cells (Stefano et al., 2014). Apart from C/EBPα, also C/EBPβ,ε and δ have been demonstrated to possess differentiation and transdifferentiation potential (Cirovic et al., 2017).

In vivo, C/EBPα is expressed at high levels in liver, adipose tissue, skin, liver, lung, peripheral-blood mononuclear cells, placenta and adrenal gland. Knockout mice have no mature granulocytes and display lung, adipocyte and hepatocyte abnormalities that are accompanied by perinatal lethal metabolic defects (Ramji and Foka, 2002; Wang et al., 1995; Zhang et al., 1997). The importance of C/EBPα in haematopoiesis was further demonstrated by conditional knockout experiments in mice. Loss of C/EBPα leads to deregulated haematopoietic stem cell functions, a block of differentiation at the myeloid commitment stage and a lack of granulocyte/monocyte progenitors (Zhang et al., 2004). Mutations of the CEBPA gene are found in around 15% of acute myeloid leukaemia (AML) cases (Lin et al., 2005); AML of different aetiology frequently shows down-regulation of CEBPA expression (Avellino and Delwel, 2017; Pabst and Mueller, 2009). The majority of CEBPA mutations are either located in the bZIP domain or are frame shift- and stop-mutations in the 5’ region of C/EBPα, that lead to enhanced expression of the P30 isoform (Fasan et al., 2014; Lin et al., 2005). Concordantly, p42-C/EBPα deficient mice engineered to express p30-C/EBPα from the Cebpa locus develop an AML-type of disease with complete penetrance (Kirstetter et al., 2008). Molecular analysis of haematopoietic cells from p30 mice revealed that p30-C/EBPα facilitated development of committed myeloid progenitors with an increased proliferation phenotype (Kirstetter et al., 2008; Bereshchenko et al., 2009).

The biological role of C/EBPα and other C/EBP transcription factors is directly connected to their modularity, and individual conserved regions confer overlapping, but distinct functions. This is highlighted by the fact that deletion and swapping of C/EBP regions does not abrogate, but alter C/EBP function with changes in transcriptional and

phenotypic outcome. For example, C/EBPβ deletion mutants with altered protein structure were still capable of transdifferentiating B-cells into myeloid cells albeit with differences in the resulting fractions of different myeloid sub-populations (Stoilova et al., 2013). A chimeric protein consisting of the C/EBPα N-terminus fused to the C/EBPβ bZIP domain induced granulocytic differentiation of myeloid K562 cells with similar efficiency compared to wild type (WT) C/EBPα. However, the authors noted significant differences in the gene expression pattern induced by these two proteins (Ferrari-Amorotti et al., 2010). Although structurally different, CR1 in C/EBPβ and TEIII (CR1L) in C/EBPα were both shown to interact with the SWI/SNF complex. This interaction was essential for the adipogenic differentiation potential of C/EBPα and fusing CR1 of C/EBPβ to the C/EBPα N-terminus functionally compensated for the loss of TEIII (Pedersen et al., 2001). A few years later it was demonstrated that methylation of an arginine residue within CR1 of C/EBPβ by PRTM4/CARM1 constraints the interaction with the SWI/SNF complex (Kowenz-Leutz et al., 2010). Furthermore, phosphorylation of C/EBPβ inhibited the interaction with PRTM4/CARM1, implying crosstalk between different C/EBP modifications that fine tunes C/EBP function.

C/EBP proteins are decorated by a multitude of PTMs including phosphorylation, acetylation as well as methylation (Dittmar et al., 2019; Leutz et al., 2011). Although understudied compared to phosphorylation, non-histone protein methylation has gained increased attention in the last years and emerged as an important regulator of cellular signal transduction (Biggar and Li, 2015). For example, a recent study reports that PRMT1-dependent methylation of C/EBPα promotes cell growth by blocking the interaction between C/EBPα and co-repressor HDAC3 and leading to cyclin D1 up-regulation (Li-ming et al., 2019). It is anticipated that unraveling the PTM and motif dependent interaction network of C/EBPα may help to understand how C/EBPα regulates gene expression in different cellular contexts such as the haematopoietic system.

1.3. Studying protein-protein interactions with mass spectrometry

Proteins do not exert their functions alone but within a complex network of protein interactions. Therefore, identifying protein-protein interactions (PPIs) is a central element of understanding protein functions in signal transduction and gene regulation. Mass spectrometry-based proteomics has emerged as an integral part of protein interaction studies. In the next paragraphs, I will give an overview about mass spectrometry-based proteomics in general and more specifically the application for PPI studies.

1.3.1. Mass spectrometry based proteomics

In their landmark review, Tyers and Mann refer to proteomics as “not only studying all the proteins in any given cell, but also the set of all protein isoforms and modifications, the interactions between them, the structural description of proteins and their higher-order complexes, and for that matter almost everything 'post-genomic'” (Tyers and Mann, 2003). The massive technological leaps forward in mass spectrometry that have occurred in the last three decades not only allow the identification but most importantly, also the simultaneous quantification of several thousands of proteins in a single sample.

By definition, mass spectrometry (MS) is an analytical technique that measures the mass to charge ratio of ions - therefore ionisation is a prerequisite for the analysis. Although electrospray ionisation is capable of ionising intact proteins (Tipton et al., 2011), digestion of proteins into peptides offers significant analytical benefits by facilitating better chromatographic separation, ionisation and interpretation of less complex MS spectra (Zhang et al., 2013). This “bottom up” or “shotgun” approach is used by the majority of proteomics labs these days and requires treatment of protein samples with a protease. Typically trypsin (cuts C-terminal of arginine and lysine) alone or in combination with LysC (cuts C-terminal of lysine) is used, but also other amino acid specific proteases are available, depending on the needs of the experiment. On-line physical separation of the peptide mixture via reversed-phase liquid chromatography (RP-LC) reduces complexity of the sample before ionisation and injection into the mass spectrometer. Different mass analysers exist to determine the mass to charge ratio of the positively ionised peptides (Zubarev and Makarov, 2013). In data-dependent acquisition mode the top N most abundant ion species, that are eluting from the chromatography at a given time point, are selected for fragmentation and analysis of the fragments. The first mass scan is referred to as MS1 scan while the second fragmentation scan is termed MS/MS or MS2 (Bozorgzadeh et al., 1978).

During fragmentation, the peptide breaks at its amide bonds, producing ions that correspond to sequence fragments of the isolated peptide. The peptide sequence can be inferred from its fragmentation spectrum and the identified peptides are assembled back into proteins by search algorithms (Hunt et al., 1981; Yates, 1998; Zhang et al., 2013).

Early proteomic efforts were focused on identifying as many proteins as possible, but strategies for reliable quantification of proteins were soon developed (Aebersold and Mann, 2003). Depending on the experimental setup, different techniques are available that in some cases require the incorporation of isotopic labels into proteins or peptides (Bantscheff et al., 2012). Stable isotope labelling by amino acids in cell culture (SILAC) has been proven to be a very useful and robust method for the labelling of proteins in cell culture (Ong et al., 2002). The basis for this labelling technique is the replacement of essential amino acids with amino acids containing naturally occurring isotopes (heavy amino acids) in the cell culture media. The differentially labeled cells (heavy, medium, light) can then be combined prior to lysis and processed together, facilitating accurate relative quantification of up to three different samples at once. Label free quantification (LFQ) on the other hand does not require the incorporation of metabolic or chemical labels into the sample but is based on the obtained raw peptide intensities that are combined into protein intensities with a complex normalisation strategy (Wu et al., 2015). A major advantage of LFQ is that no extra time or cost intensive labelling is required and that more than three different samples can be compared at once. However, due to the higher variability in sample preparation (samples are not processed together as in SILAC), more replicates are required and small changes are harder to quantify (Wu et al., 2015).

1.3.2. Protein interaction studies

With sound quantification strategies at hand, mass-spectrometry has become the method of choice for identifying and quantifying PPIs. There are different ways to capture or pull-down protein interactors of a protein of interest (bait). In **Figure 4,** an overview of three major pull-down strategies is presented. The enriched interactors from the pull-down workflow are then subsequently analysed with shotgun mass spectrometry.

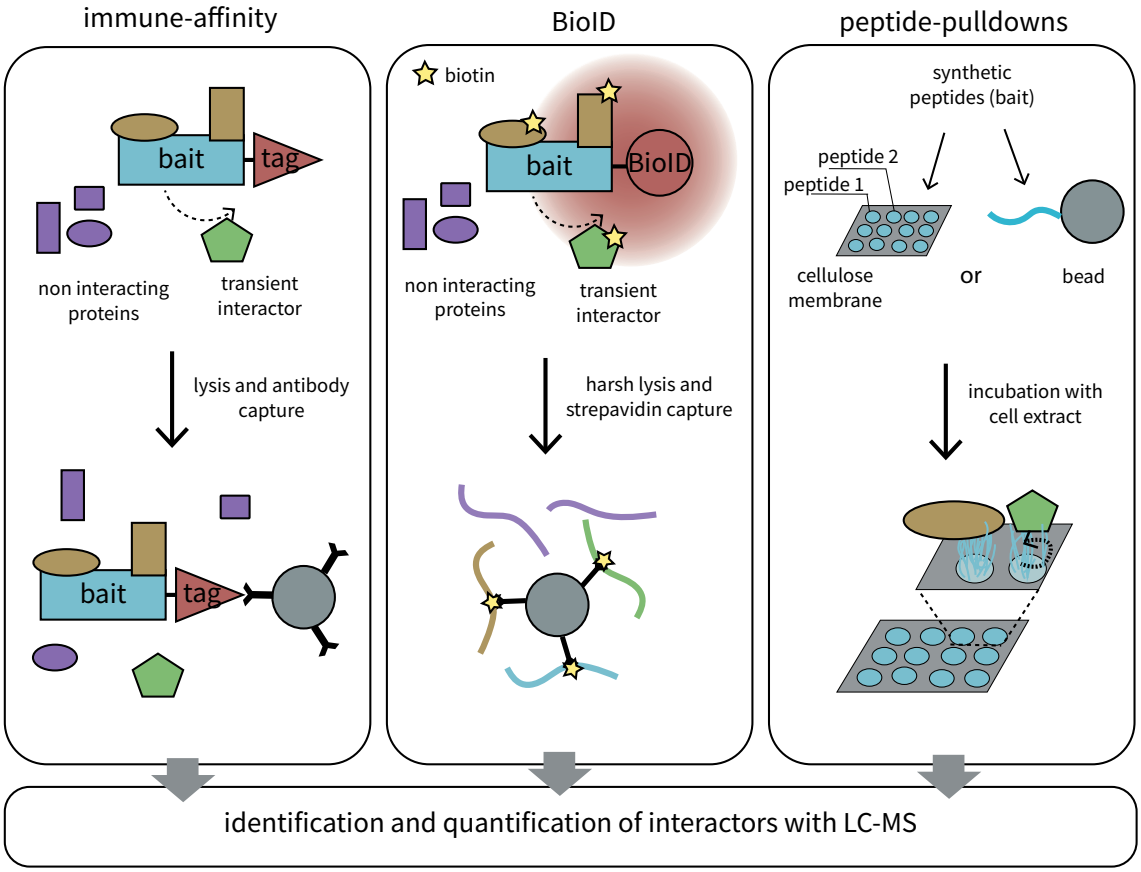


Figure 4: Schematic representation of different pull-down strategies.
In immune-affinity pull-downs, the bait and interactors are enriched with an antibody directed either against the bait itself or an affinity tag. In BioID experiments, a promiscuous biotin ligase covalently attaches biotin to proximal proteins. In peptide pull-downs, synthetic peptides coupled to beads or a membrane support are screened for protein interactions. The enriched proteins in the pull-down can subsequently be identified and quantified with shotgun proteomics.

1.3.2.1. AP-MS

In affinity purification coupled to LC-MS (AP-MS), the bait is purified from a cell lysate with an antibody directed against either the bait protein itself or an affinity tag that has been genetically fused to the bait (Dunham et al., 2012). Interacting proteins are co-purified together with the bait while background proteins are washed away in subsequent washing steps. In the early days of AP-MS, pull-downs were often sequentially purified with two different antibodies in order to reduce the amount of background binders (tandem affinity purification). Remaining proteins were subsequently identified with non-quantitative MS (Puig et al., 2001). With the rapid improvement of mass spectrometry technologies, quantitative MS now aids the discrimination of interactors from contaminating background proteins (Keilhauer et al., 2014; Meyer and Selbach, 2015). Typically, negative control samples are included in the experiment to discriminate background from true interactors. However, even with these technical improvements, AP-MS has some limitations when it comes to detecting PPIs. If proteins are purified through an affinity tag, genetic engineering is necessary and overexpression of the bait may lead to false positives. During lysis, proteins residing in different compartments in the cell that naturally don't interact are brought into proximity of each other, and might form interactions during the experimental procedure. Additionally, protein interactions that are weak or transient are easily lost during the purification process. Interactions mediated by IDPs that are implicated in gene regulation and signalling events frequently fall into these categories. AP-MS experiments using the whole protein as bait do not give information about which part of the protein is mediating the interaction. Furthermore, detecting the influence of PTMs on PPIs is challenging with traditional AP-MS workflows.

1.3.2.2. Proximity labelling

Proximity labelling of interactors with biotinylating enzymes has become increasingly popular over the last years and can in part overcome some of the issues of AP-MS workflows. Roux et al. have published a method in 2012 named BioID (proximity-dependent biotin identification) that employs a promiscuous biotin ligase fused to the protein of interest (Roux et al., 2012, 2018). The enzyme (BirA*, also referred to as BioID) is a mutated version of the E*scherichia coli* biotin ligase BirA and capable of converting biotin into highly reactive biotinyl-5'-AMP (bioAMP) that reacts with lysine residues in close proximity. Current estimations suggest 10 to 50 nm labelling radius, although this number also depends on flexible linker regions that can be included in the construct (Trinkle-Mulcahy, 2019). Biotinylated proteins can then be

enriched via the extremely high affinity interaction with streptavidin (Kd = 10^{-14}) (Green, 1963). An improved biotin ligase that is smaller and faster (BioID2) than the original BioID version was soon published after the original BioID paper (Kim et al., 2016). Most recent developments include an even faster version of the enzyme (TurboID) that reduces the required labelling time form originally 24 hours to only 10 minutes (Branon et al., 2018). A clear advantage of proximity labelling over AP-MS is that the covalent biotin modification stays attached to the interactor even when the interactor dissociates. This facilitates recovery of transient or weak interactions. Additionally, the high affinity between streptavidin and biotin allows for stringent washing steps and significant background reduction (Roux et al., 2018).

Similar to AP-MS experiments with an affinity tag, BioID experiments require fusing the ligase to the bait and introducing the transgene into the model system of choice. Biotin easily diffuses through cell membranes and is added to the cell culture media for the duration of the labelling. An alternative enzyme for biotin proximity labelling is APEX, an engineered enzyme derived from soy or pea ascorbate peroxidase (Martell et al., 2012; Rhee et al., 2013). APEX requires treatment of the specimen with biotin phenol and hydrogen peroxide, creating biotin–phenol radicals that react with proteins in their vicinity. While APEX has only been used by a handful of labs, BioID has already been widely applied in over 100 different studies. Care should be taken that the BioID tag, which is a bit larger than GFP in size, does not interfere with location, function or stability of the bait protein. In BioID experiments proteins proximal to the bait are identified, which does not necessarily indicate a direct interaction. On the other hand, the lack of accessible primary amines for biotinylation in a direct interactor may lead to false negative results.

1.3.2.3. Peptide-protein pull-downs

PPIs mediated by short, unstructured amino acid sequences are implicated in signal transduction and gene regulation (Wright and Dyson, 2015). These PPIs can be recapitulated by peptide-protein pull-down assays that employ synthetic peptides as baits. In chemical peptide synthesis not only naturally occurring amino acid, but also modification carrying amino acids – e.g. phosphorylated, acetylated or methylated amino acids – can be incorporated. This aids the detection of PTM specific binding events and poses a clear advantage of peptide-protein pull-downs (Schulze and Mann, 2004; Schulze et al., 2005; Tinti et al., 2014; Dittmar et al., 2019). Capture of interacting proteins is typically facilitated by immobilizing the synthetic peptides on beads through a linker group (Lange et al., 2010; Schulze and Mann, 2004; Schulze et al., 2005;

Selbach et al., 2009) or by using peptide libraries synthesised on a solid membrane support (Frank and Overwin, 1996; Lachner et al., 2001; Wiedemann et al., 2004). The introduction of SPOT synthesis in 1992 (Frank, 1992) greatly facilitated and reduced costs of synthetic peptide array preparation.

In the past, peptide arrays have been widely used by numerous studies for epitope mapping (Gao and Esnouf, 1996; Forsström et al., 2014; Reineke and Sabat, 2008) and mapping of protein interactions with an antibody based approach similar to far western blotting (Katz et al., 2011; Volkmer et al., 2012). While the latter is focused on the detection of a specific protein by an antibody coupled to imaging techniques, mass spectrometry aids the unbiased identification of interacting proteins. In the peptide array X-linking (PAX) assay, synthetic peptide arrays are incubated with cell lysate, followed by crosslinking of the interactors. After washing, peptide spots are excised and prepared for analysis with mass spectrometry (Okada et al., 2012).

A similar workflow without crosslinking was implemented in a Protein Interaction Screen on a peptide Matrix (PRISMA) that mapped protein interactions to the amino acid sequence of C/EBPβ (Dittmar et al., 2019). In detail, 14 amino acid long peptides, designed with a sequence overlap of four amino acids (tiling peptides), were synthesised on a cellulose membrane and probed for protein interactions with nuclear cell lysate. The authors included 201 C/EBPβ derived peptides with and without PTMs in the screen and detected interaction footprints for over 1000 proteins across the C/EBPβ sequence and PTM sites. A similar screen by Meyer et al. employed peptide array pull-downs to detect the impact of disease causing point mutations on protein interactions (Meyer et al., 2018). An advantage of the PRISMA method or of cellulose peptide membranes in general is the high local density of peptides that can be achieved. According to the manufacturer (JPT, Berlin, Germany), the peptide arrays that were used by Meyer at al., 2018 and Dittmar et al.,2019, carried 5 nmol of peptide per spot which would translate to an approximate peptide density of 520 nmol/cm^2. High local peptide concentrations may counteract the dissociation of transient interactors by providing a second binding site after dissociation of the weak interactor (Ruthenburg et al., 2007).

1.4. Aim of this study

PPIs may be part of the puzzle that explains the modularity and plasticity of C/EBPα functions. In the past, several attempts have been made to catalogue the C/EBPα interactome with AP-MS based studies (Cirilli et al., 2017; Giambruno et al., 2013; Grebien et al., 2016). However, the disordered and PTM decorated structure of C/EBPα poses a significant challenge for antibody-based pull-downs and low overlaps (0 to 5%) between C/EBPα interactomes generated with different AP-MS workflows have been reported (Giambruno et al., 2013). For many known C/EBPα interactors it is unclear which region of C/EBPα mediates the interaction and to which extent PTMs are involved. The aim of this study is to apply PRISMA and BioID to comprehensively map the C/EBPα interactome across conserved regions and PTM sites. The isoform-specific interactome is of particular interest as the two C/EBPα isoforms have different biological functions in the cell and the P30 proteoform acts as an oncogene in AML. In this thesis, the interactome will be used as a proxy to annotate functionality of individual conserved regions in C/EBPα.

In this context, another focus of the present work is to detect novel modification sites of C/EBPα and elucidate their impact on protein interactions. A number of PTMs have been described for C/EBPα and the amino acid sequence is rich in conserved arginines and lysines that are potential targets of methylating enzymes.

Protein interactions may be context- and cell-specific. Considering the importance of C/EBPα in granulocytic differentiation, I chose the myeloid NB4 cell line as a model system for C/EBPα protein interaction studies. NB4 cells can be induced to differentiate into granulocytes or macrophages with all-trans retinoic acid (ATRA) or 12-O-Tetradecanoylphorbol-13-acetate (TPA), respectively. In the present work, I describe transcriptomic and proteomic changes occurring during this differentiation process and evaluate the role of C/EBPα in this context.

2. Materials and Methods

2.1. Cell culture

NB4 cells were acquired from Leibniz Institute DSMZ- German Collection of Microorganisms and Cell Culture, Germany (DSMZ no.: ACC 207). Cells were cultivated in a humidified incubator at 37°C, 5% CO2 in RPMI1640 supplemented with, 1x GlutaMAX, 10% FCS and 100mg/ml penicillin streptomycin (all from Gibco™, Thermo Fisher Scientific, Germany).

2.2. SILAC labelling of NB4 cells

For metabolic labelling, NB4 cells were grown in SILAC RPMI1640 supplemented with 10% dialyzed FCS, 100 mg/ml penicillin-streptomycin, 25mM HEPES, 28 µg/ml L-arginine and 48.67 µg/ml L-lysine $^{13}C_6{}^{15}N_2$ (heavy lysine) or L-lysine D4 (medium lysine). Complete labelling of proteins was confirmed prior to experiments. Media and supplements for SILAC experiments were purchased from Thermo Fisher Scientific, Germany.

2.3. NB4 differentiation

NB4 were seeded in a 6-well plate at a density of 0.5x 10^6/ml in SILAC media supplemented with 2µM ATRA, 50nM TPA or solvent control (0.0012% DMSO), all purchased from Sigma-Aldrich, Germany. After two days, cells differentiated with ATRA were diluted 1:2 with fresh media supplemented with ATRA and after 4 days the media was exchanged with fresh media supplemented with ATRA.

2.4. Surface marker staining and FACS analysis

Cells treated with ATRA, solvent control and TPA time-points < 16 h of treatment were harvested by centrifugation and washed once with ice cold PBS (Gibco™, Thermo Fisher Scientific, Germany). Adherent cells (TPA time points 16h and later) were washed once with ice cold PBS and harvested by trypsinisation. After washing, 250000 cells were stained with PE Mouse Anti Human CD11b antibody (BD-Pharmingen, Clone ICRF44) diluted 1:25 in FACS buffer (PBS, 1% FCS) for 30 min in the dark on ice. The cells were washed once with ice-cold FACS buffer and resuspended in FACS buffer for analysis. Data was acquired on a BD LSRII flow cytometer, recording 10000 events per sample. FACS data was analysed with FlowJo software and the gate of CD11b positive cells was adjusted to unstained differentiated cells and stained control cells.

2.5. Wright Giemsa staining

Cells were stained with May-Grünewald and Giemsa staining as described before (Cirovic et al., 2017). In detail, 75000 cells per sample were collected on a glass slide with a cytospin centrifuge (5 min, 500g) and air-dried. Slides were immersed in May Grünewald stain for 5 min, followed by rinsing with PBS and staining with Giemsa stain diluted 1:20 with ddH_2O for 20 min. Slides were rinsed once more with PBS, air-dried and mounted.

2.6. RNA extraction

Total RNA was extracted using RNA NOW™ reagent (Ozyme, France) following manufacturer's instructions. In brief, cell pellets were homogenised by resuspending in 750µl of RNA NOW reagent per tube, followed by adding 200µl of chloroform. Samples were shaken by hand, incubated for 5 min on ice and centrifuged at 4°C for 5 min at 16000g. The upper phase was transferred into a new tube and one volume of isopropanol was added. The samples were incubated for 1h at -20°C and subsequently centrifuged at 4°C for 1h at 16000g. The resulting pellet was washed twice with 75% ethanol and dissolved in 30µl RNAse free H_2O. RNA was quantified and 10µg RNA was prepared in 50µl RNAse free H_2O. DNA was removed with a DNA removal kit (DNA-free Kit, Thermo Fisher Scientific). DNA was digested by adding 5µl of 10x DNAseI buffer and 1µl of DNAseI and incubation at 37°C for 30 minutes. After incubation 5µl of DNAse inactivation reagent was added and the sample was incubated for 2 minutes at RT, mixing occasionally by hand. The samples were centrifuged at RT for 1.5 min at 1000g and the RNA was transferred into a new tube. RNA concentration was measured by Nanodrop and the integrity of purified RNA was checked with a bioanalyzer chip. Only RNA samples with a RNA integrity score (RIN) >7.6 were processed for further analysis.

2.7. Microarrays

Microarray analysis and raw data processing was performed in the Genome Research Unit at the Luxembourg Institute of Health by Nathalie Nicot, Petr Nazarov and Arnaud Muller. RNA expression was analysed with ClariomS human assays (ThermoFisher Scientific) covering 20000 annotated human genes (no isoforms). The raw microarray CEL files were imported into Transcriptome Analysis Console software of ThermoFisher. The Robust Multichip Average with GC correction (SST RMA) method was applied to the data set resulting in expression values for transcript clusters. To decrease number of uninformative features, only transcript clusters with log_2

expression above 6 in at least one sample were considered for further analysis. The differentially expressed genes were identified using limma R/Bioconductor package (Ritchie et al., 2015) with Benjamini-Hochberg's FDR correction for multiple testing.

2.8. Whole cell protein extract preparation

Whole cell protein extracts were prepared by methanol chloroform extraction as described previously (Sapcariu et al., 2014). In brief, suspension cells were harvested via centrifugation and washed twice with ice-cold PBS. Proteins were extracted by adding equal volumes (400µl for 6 well plate) of ice-cold methanol, water and chloroform to the cell pellets. Samples were agitated for 20 min on a tube shaker at 4°C at 1400rpm, followed by centrifugation at 4°C for 10 min at 18000g. The resulting upper phase was removed and the interphase containing the proteins was washed with 1ml ice-cold methanol. The methanol was removed and the pellet air-dried and resuspended in 100µl denaturation buffer (6M urea, 2M thiourea, 10mM HEPES, pH 8.0, all from Sigma-Aldrich, Germany). The sample was sonicated with 5 pulses with a probe sonicator on ice and centrifuged at 4°C for 20 min at 18000g. The supernatant was transferred to a fresh tube and protein concentration was determined with a Bradford protein assay. Proteins were digested into peptides by in solution digestion as described below.

2.9. Nuclear extract preparation

Nuclear extracts from NB4 cells were prepared as described previously (Dignam et al., 1983) with slight modifications. NB4 cells were harvested by centrifugation at 4°C for 10 min at 1000g and washed twice with ice cold PBS. Packed cell volume (pcv) was estimated and cells were resuspended in 5x pcv of ice-cold hypotonic buffer (10mM HEPES pH 7.5, 10mM NaCl, 3mM MgCl2) supplemented with protease inhibitors. Cells were incubated on ice for 5 min, followed by addition of n-Dodecyl β-D-maltoside (Sigma-Aldrich, Germany) to a final concentration of 0.02% from a 10% stock solution. The sample was vortexed for 2s and immediately centrifuged at 4°C for 5 min at 600g. The cytosolic fraction was removed and the nuclei were washed with 20x pcv hypotonic buffer (5 min, 600g, 4°C). The supernatant was removed and the nuclei were washed with 20x pcv PBS (4°C, 5 min, 600g). The nuclei were extracted with 2/3x pcv of high salt buffer (20mM HEPES pH 7.5, 400mM NaCl, 1mM EDTA pH 8, 1mM EGTA pH 8, 20% glycerol, 1mM DTT) supplemented with protease inhibitors (cOmplete Mini, EDTA-free, Sigma-Aldrich, Germany) while shaking on a tubeshaker at 4°C for 20 min at 750 rpm. Nuclear extracts were cleared by

centrifugation at 4°C for 20 min at 18000g and the buffer was exchanged by gel filtration with PD MidiTrap G10 columns (GE healthcare) according to manufacturers instructions. In brief, columns were equilibrated three times with 5ml membrane binding buffer (20mM HEPES pH 7.5, 400mM NaCl, 1mM EDTA pH 8, 1mM EGTA pH 8, 25% glycerol, 1mM DTT) and the flow through was discarded. The sample was loaded on the column and centrifuged at 4°C for 2 min at 1000g, collecting the eluate (nuclear extract in membrane binding buffer) in a fresh tube. Nuclear extracts were snap frozen in liquid nitrogen. Protein concentration of nuclear extracts ranged between 6 and 7 mg/ml.

2.10. Determination of protein concentration with a Bradford assay

Concentration of protein extracts was determined with a Bradford assay (Bradford, 1976). Protein extracts were diluted 1:10 with H_2O and 2µl of sample were mixed with 498µl of H_2O and 500µl of Coomassie Bradford Protein Assay reagent (PierceTM, Thermo Fisher Scientific, Germany) in a cuvet. The sample was vortexed and after 5 min incubation at RT the absorbance at 595 nm was measured. Protein concentration of the sample was inferred from a serial dilution of a bovine serum albumin (BSA) standard spanning 0.1 to 2 mg/ml.

2.11. Western blotting

Protein extracts (15µg protein/sample) were mixed with 4x loading buffer (Bio-Rad) and boiled for 5 min at 95°C. Samples were loaded on a precast 10-12% SDS-polyacrylamide gel (Protean, Bio-Rad) and separated by electrophoresis at 120V in running buffer (25mM Tris, 200mM Glycine, 0.1% SDS). Proteins were transferred to a nitrocellulose membrane with the Trans-Blot Turbo Midi System from Bio-Rad and successful transfer was confirmed by staining the membrane with Ponceau S solution for 5 min. Free binding sites on the membrane were blocked by incubation with 5% skimmed milk in TBS-T (50mM Tris-HCl, 150mM NaCl, 0.1% Tween-20) at RT for 1h. For detection of biotinylation, membranes were incubated with streptavidin-HRP (Sigma-Aldrich, Germany) diluted 1:2000 in blocking solution at RT for 1h. For detection of actin, histone H3 or flag expression, membranes were incubated with primary antibodies diluted in blocking solution overnight at 4°C. Antibodies were purchased from Abcam (histone H3: ab1791, actin: ab179467, Flag: ab49763). Membranes were washed 3x for 5 min in TBS-T and incubated 1h at RT with an HRP-coupled secondary antibody raised against the species of the primary antibody. Membranes were washed 3x for 5 min in TBS-T and immersed in chemiluminescence

reaction solution (Milipore) for 1 min. Bands were detected using the C-DiGit Blot Scanner.

2.12. Stable NB4 cell lines

NB4 stable cell lines used in this thesis were generated by Dr. Elisabeth Kowenz-Leutz and Valeria Sapozhnikova. In brief, the sequence of rat p42-Cebpa, p30-Cebpa or p30-Cebpa-3L-mutant (R140, 147, 154 -> L) was C-terminally fused to a promiscuous biotin ligase (Roux et al., 2012) containing a C-terminal FLAG-tag. A flexible GS-linker was inserted between Cebpa and BioID. Biotin ligase is referred to in this thesis as BioID, Cebpa-BioID fusions are referred to as C/EBPα-BioID. BioID and C/EBPα-BioID fusion constructs were cloned into the inducible lentiviral pInducer21 gene expression vector containing an IRES eGFP marker (Meerbrey et al., 2011). Following infection, successfully transduced NB4 cells were selected by FACS sorting for GFP fluorescence (> 98 % GFP positive cells). NB4 stable cell lines were maintained in RPMI media supplemented with tetracycline FCS (Sigma Aldrich, Germany). Culture conditions were the same as for parental NB4 cells as described in section 2.1.

2.13. Protein Interaction Screen on a peptide Matrix

Protein interaction screen on a peptide matrix was performed as described before (Dittmar et al., 2019) with slight adaptations. Custom PepSpot cellulose membranes were ordered from JPT (Berlin, Germany). C/EBPα peptides contained on the peptide array are summarised in Table 1. All washing and incubation steps were performed on a rocking platform set to 700 rpm. Prior to the experiment, membranes were conditioned with membrane binding buffer (20mM HEPES pH 7.5, 400mM NaCl, 1mM EDTA pH 8, 1mM EGTA pH 8, 25% glycerol, 1mM DTT) for 15 min at room temperature, followed by a blocking step with 1 mg/ml yeast tRNA (Sigma-Aldrich) in membrane binding buffer for 10 min at RT. Membranes were washed 5 x for 5 min with membrane binding buffer and then placed into a polypropylene bag. SILAC labeled nuclear extracts from NB4 cells (H/M/L) were mixed just before incubation (final protein concentration 6mg/ml) and slowly added into the polypropylene bag. The bag was sealed and placed on ice on the rocking platform. The membranes were incubated for 30 min, followed by two washing steps (5 min each) with ice-cold membrane binding buffer and one washing step (5 min) with ice-cold membrane binding buffer without glycerol. The membranes were placed on a glass slide and air-dried. The individual

peptide spots were punched out with a 3 mm biopsy puncher (Stiefel, Germany) and placed into single wells of a 96 well plate containing 20µl denaturation buffer (6M urea, 2M thiourea, 10mM HEPES pH 8). Samples were digested (in solution digestion protocol), desalted and analysed as described in the following sections.

peptide sequence	unmodified sequence	ID	region	AA start	AA end	mod-ification	Pearson correlation replicates
MESADFYEAEPRPPM	MESADFYEAEPRPPM	1	CR2	1	15	[]	0.90
M-Nterm.ac-ESADFYEAEPRPPM	MESADFYEAEPRPPM	2	CR2	1	15	1', 'N-term ac',	0.88
MESADFYEAEPR-me2_sym-PPM	MESADFYEAEPRPPM	3	CR2	1	15	12', 'me2_sym',	0.90
MESADFYEAEPR-me2_asym-PPM	MESADFYEAEPRPPM	4	CR2	1	15	12', 'me2_asym',	0.81
EAEPRPPMSSHLQSP	EAEPRPPMSSHLQSP	5		8	22	[]	0.75
EAEPR-me2_sym-PPMSSHLQSP	EAEPRPPMSSHLQSP	6		8	22	12', 'me2_sym',	0.76
EAEPRPPMSSHLQS-phos-P	EAEPRPPMSSHLQSP	7		8	22	21', 'phos',	0.62
EAEPR-me2_asym-PPMSSHLQSP	EAEPRPPMSSHLQSP	8		8	22	12', 'me2_asym',	0.74
MSSHLQSPPHAPSSA	MSSHLQSPPHAPSSA	9		15	29	[]	0.85
MSSHLQS-phos-PPHAPSSA	MSSHLQSPPHAPSSA	10		15	29	21', 'phos',	0.85
PPHAPSSAAFGFPRG	PPHAPSSAAFGFPRG	11		22	36	[]	0.91
PPHAPSSAAFGFPR-me2_sym-G	PPHAPSSAAFGFPRG	12		22	36	35', 'me2_sym',	0.91
PPHAPSSAAFGFPR-me2_asym-G	PPHAPSSAAFGFPRG	13		22	36	35', 'me2_asym',	0.89
AAFGFPRGAGPAQPP	AAFGFPRGAGPAQPP	14		29	43	[]	0.87
AAFGFPR-me2_sym-GAGPAQPP	AAFGFPRGAGPAQPP	15		29	43	35', 'me2_sym',	0.88
AAFGFPR-me2_asym-GAGPAQPP	AAFGFPRGAGPAQPP	16		29	43	35', 'me2_asym',	0.93
GAGPAQPPAPPAAPE	GAGPAQPPAPPAAPE	17		36	50	[]	0.72
PAPPAAPEPLGGICE	PAPPAAPEPLGGICE	18	CR3	43	57	[]	0.65
EPLGGICEHETSIDI	EPLGGICEHETSIDI	19	CR3	50	64	[]	0.77
EHETSIDISAYIDPA	EHETSIDISAYIDPA	20	CR3	57	71	[]	0.71
ISAYIDPAAFNDEFL	ISAYIDPAAFNDEFL	21	CR3	64	78	[]	0.92
AAFNDEFLADLFQHS	AAFNDEFLADLFQHS	22	CR4	71	85	[]	0.93
LADLFQHSRQQEKAK	LADLFQHSRQQEKAK	23	CR4	78	92	[]	0.89
LADLFQHSRQQEK-ac-AK	LADLFQHSRQQEKAK	24	CR4	78	92	90', 'ac',	0.86
LADLFQHSR-me2_asym-QQEKAK	LADLFQHSRQQEKAK	25	CR4	78	92	86', 'me2_asym',	0.83
LADLFQHSR-me2_sym-QQEKAK	LADLFQHSRQQEKAK	26	CR4	78	92	86', 'me2_sym',	0.81

LADLFQHSRQQEK-me2-AK	LADLFQHSRQQEKAK	27	CR4	78	92	90', 'me2',	0.85
SRQQEKAKAAVGPTG	SRQQEKAKAAVGPTG	28		85	99	[]	0.90
SRQQEKAK-me2-AAVGPTG	SRQQEKAKAAVGPTG	29		85	99	92', 'me2',	0.90
SR-me2_asym-QQEKAKAAVGPTG	SRQQEKAKAAVGPTG	30		85	99	86', 'me2_asym',	0.90
SR-me2_sym-QQEKAKAAVGPTG	SRQQEKAKAAVGPTG	31		85	99	86', 'me2_sym',	0.90
SRQQEK-ac-AKAAVGPTG	SRQQEKAKAAVGPTG	32		85	99	90', 'ac',	0.86
SRQQEKAK-ac-AAVGPTG	SRQQEKAKAAVGPTG	33		85	99	92', 'ac',	0.87
SRQQEK-me2-AKAAVGPTG	SRQQEKAKAAVGPTG	34		85	99	90', 'me2',	0.90
KAAVGPTGGGGGGDF	KAAVGPTGGGGGGDF	35		92	106	[]	0.46
K-me2-AAVGPTGGGGGGDF	KAAVGPTGGGGGGDF	36		92	106	92', 'me2',	0.55
K-ac-AAVGPTGGGGGGDF	KAAVGPTGGGGGGDF	37		92	106	92', 'ac',	0.80
GGGGGGDFDYPGAPA	GGGGGGDFDYPGAPA	38	CR5	99	113	[]	0.78
FDYPGAPAGPGGAVM	FDYPGAPAGPGGAVM	39	CR5	106	120	[]	0.81
AGPGGAVMPGGAHGP	AGPGGAVMPGGAHGP	40		113	127	[]	0.89
MPGGAHGPPPGYGCA	MPGGAHGPPPGYGCA	41	CR1L	120	134	[]	0.82
M-Nterm.ac-PGGAHGPPPGYGCA	MPGGAHGPPPGYGCA	42	CR1L	120	134	['120', 'N-term ac']	0.83
PPPGYGCAAAGYLDG	PPPGYGCAAAGYLDG	43	CR1L	127	141	[]	0.78
AAAGYLDGRLEPLYE	AAAGYLDGRLEPLYE	44	CR1L	134	148	[]	0.79
AAAGYLDGR-me2_asym-LEPLYE	AAAGYLDGRLEPLYE	45	CR1L	134	148	142', 'me2_asym',	0.82
AAAGYLDGR-me2_sym-LEPLYE	AAAGYLDGRLEPLYE	46	CR1L	134	148	142', 'me2_sym',	0.88
GRLEPLYERVGAPAL	GRLEPLYERVGAPAL	47	CR1L	141	155	[]	0.90
GR-me2_asym-LEPLYERVGAPAL	GRLEPLYERVGAPAL	48	CR1L	141	155	142', 'me2_asym',	0.90
GR-me2_sym-LEPLYERVGAPAL	GRLEPLYERVGAPAL	49	CR1L	141	155	142', 'me2_sym',	0.88
GRLEPLYER-me2_asym-VGAPAL	GRLEPLYERVGAPAL	50	CR1L	141	155	149', 'me2_asym',	0.86
GRLEPLYER-me2_sym-VGAPAL	GRLEPLYERVGAPAL	51	CR1L	141	155	149', 'me2_sym',	0.91
ERVGAPALRPLVIKQ	ERVGAPALRPLVIKQ	52	CR1L	148	162	[]	0.77
ER-me2_asym-VGAPALRPLVIKQ	ERVGAPALRPLVIKQ	53	CR1L	148	162	149', 'me2_asym',	0.87
ERVGAPALR-me2_sym-PLVIKQ	ERVGAPALRPLVIKQ	54	CR1L	148	162	156', 'me2_sym',	0.85
ERVGAPALR-me2_asym-PLVIKQ	ERVGAPALRPLVIKQ	55	CR1L	148	162	156', 'me2_asym',	0.83
ERVGAPALRPLVIK-me2-Q	ERVGAPALRPLVIKQ	56	CR1L	148	162	161', 'me2',	0.83
ERVGAPALRPLVIK-ac-Q	ERVGAPALRPLVIKQ	57	CR1L	148	162	161', 'ac',	0.89
ER-me2_sym-VGAPALRPLVIKQ	ERVGAPALRPLVIKQ	58	CR1L	148	162	149', 'me2_sym',	0.59

LRPLVIKQEPREEDE	LRPLVIKQEPREEDE	59	CR6	155	169	[]	0.78
LR-me2_sym-PLVIKQEPREEDE	LRPLVIKQEPREEDE	60	CR6	155	169	156', 'me2_sym',	0.64
LR-me2_asym-PLVIKQEPREEDE	LRPLVIKQEPREEDE	61	CR6	155	169	156', 'me2_asym',	0.74
LRPLVIKQEPR-me2_sym-EEDE	LRPLVIKQEPREEDE	62	CR6	155	169	164', 'me2_sym',	0.76
LRPLVIK-me2-QEPREEDE	LRPLVIKQEPREEDE	63	CR6	155	169	161', 'me2',	0.87
LRPLVIK-ac-QEPREEDE	LRPLVIKQEPREEDE	64	CR6	155	169	161', 'ac',	0.83
LRPLVIKQEPR-me2_asym-EEDE	LRPLVIKQEPREEDE	65	CR6	155	169	164', 'me2_asym',	0.72
LRPLVIKQEPR-citr-EEDE	LRPLVIKQEPREEDE	66	CR6	155	169	164', 'citr',	0.84
QEPREEDEAKQLALA	QEPREEDEAKQLALA	67	CR6	162	176	[]	0.56
QEPREEDEAK-me2-QLALA	QEPREEDEAKQLALA	68	CR6	162	176	171', 'me2',	0.61
QEPREEDEAK-ac-QLALA	QEPREEDEAKQLALA	69	CR6	162	176	171', 'ac',	0.79
QEPR-me2_asym-EEDEAKQLALA	QEPREEDEAKQLALA	70	CR6	162	176	164', 'me2_asym',	0.72
QEPR-citr-EEDEAKQLALA	QEPREEDEAKQLALA	71	CR6	162	176	164', 'citr',	0.71
QEPR-me2_sym-EEDEAKQLALA	QEPREEDEAKQLALA	72	CR6	162	176	164', 'me2_sym'	0.59
EAKQLALAGLFPYQP	EAKQLALAGLFPYQP	73		169	183	[]	0.85
EAK-ac-QLALAGLFPYQP	EAKQLALAGLFPYQP	74		169	183	171', 'ac',	0.85
EAK-me2-QLALAGLFPYQP	EAKQLALAGLFPYQP	75		169	183	171', 'me2',	0.85
AGLFPYQPPPPPPPS	AGLFPYQPPPPPPPS	76		176	190	[]	0.89
PPPPPPPSHPHPHPP	PPPPPPPSHPHPHPP	77		183	197	[]	0.91
PPPPPPPS-phos-HPHPHPP	PPPPPPPSHPHPHPP	78		183	197	190', 'phos',	0.90
SHPHPHPPPAHLAAP	SHPHPHPPPAHLAAP	79		190	204	[]	0.82
S-phos-HPHPHPPPAHLAAP	SHPHPHPPPAHLAAP	80		190	204	190', 'phos',	0.84
PPAHLAAPHLQFQIA	PPAHLAAPHLQFQIA	81	CR7	197	211	[]	0.85
PHLQFQIAHCGQTTM	PHLQFQIAHCGQTTM	82	CR7	204	218	[]	0.80
AHCGQTTMHLQPGHP	AHCGQTTMHLQPGHP	83	CR7	211	225	[]	0.88
MHLQPGHPTPPPTPV	MHLQPGHPTPPPTPV	84	CR7	218	232	[]	0.86
MHLQPGHPT-phos-PPPTPV	MHLQPGHPTPPPTPV	85	CR7	218	232	226', 'phos',	0.87
MHLQPGHPTPPPT-phos-PV	MHLQPGHPTPPPTPV	86	CR7	218	232	230', 'phos',	0.84
PTPPPTPVPSPHPAP	PTPPPTPVPSPHPAP	87	CR7	225	239	[]	0.64
PT-phos-PPPTPVPSPHPAP	PTPPPTPVPSPHPAP	88	CR7	225	239	226', 'phos',	0.71
PTPPPT-phos-PVPSPHPAP	PTPPPTPVPSPHPAP	89	CR7	225	239	230', 'phos',	0.82
PTPPPTPVPS-phos-PHPAP	PTPPPTPVPSPHPAP	90	CR7	225	239	234', 'phos',	0.84

VPSPHPAPALGAAGL	VPSPHPAPALGAAGL	91		232	246	[]	0.77
VPS-phos-PHPAPALGAAGL	VPSPHPAPALGAAGL	92		232	246	234', 'phos',	0.82
PALGAAGLPGPGSAL	PALGAAGLPGPGSAL	93		239	253	[]	0.83
LPGPGSALKGLGAAH	LPGPGSALKGLGAAH	94		246	260	[]	0.84
LPGPGSALK-ac-GLGAAH	LPGPGSALKGLGAAH	95		246	260	254', 'ac',	0.77
LPGPGSALK-me2-GLGAAH	LPGPGSALKGLGAAH	96		246	260	254', 'me2',	0.87
LKGLGAAHPDLRASG	LKGLGAAHPDLRASG	97		253	267	[]	0.93
LK-me2-GLGAAHPDLRASG	LKGLGAAHPDLRASG	98		253	267	254', 'me2',	0.89
LK-ac-GLGAAHPDLRASG	LKGLGAAHPDLRASG	99		253	267	254', 'ac',	0.78
LKGLGAAHPDLRAS-phos-G	LKGLGAAHPDLRASG	100		253	267	266', 'phos',	0.89
LKGLGAAHPDLR-me2_sym-ASG	LKGLGAAHPDLRASG	101		253	267	264', 'me2_sym',	0.82
LKGLGAAHPDLR-me2_asym-ASG	LKGLGAAHPDLRASG	102		253	267	264', 'me2_asym',	0.90
HPDLRASGGSGAGKA	HPDLRASGGSGAGKA	103		260	274	[]	0.92
HPDLR-me2_asym-ASGGSGAGKA	HPDLRASGGSGAGKA	104		260	274	264', 'me2_asym',	0.86
HPDLRAS-phos-GGSGAGKA	HPDLRASGGSGAGKA	105		260	274	266', 'phos',	0.90
HPDLR-me2_sym-ASGGSGAGKA	HPDLRASGGSGAGKA	106		260	274	264', 'me2_sym',	0.91
GGSGAGKAKKSVDKN	GGSGAGKAKKSVDKN	107	bZIP	267	281	[]	0.95
AKKSVDKNSNEYRVR	AKKSVDKNSNEYRVR	108	bZIP	274	288	[]	0.97
NSNEYRVRRERNNIA	NSNEYRVRRERNNIA	109	bZIP	281	295	[]	0.95
RRERNNIAVRKSRDK	RRERNNIAVRKSRDK	110	bZIP	288	302	[]	0.95
AVRKSRDKAKQRNVE	AVRKSRDKAKQRNVE	111	bZIP	295	309	[]	0.95
KAKQRNVETQQKVLE	KAKQRNVETQQKVLE	112	bZIP	302	316	[]	0.96
ETQQKVLELTSDNDR	ETQQKVLELTSDNDR	113	bZIP	309	323	[]	0.80
ELTSDNDRLRKRVEQ	ELTSDNDRLRKRVEQ	114	bZIP	316	330	[]	0.77
RLRKRVEQLSRELDT	RLRKRVEQLSRELDT	115	bZIP	323	337	[]	0.90
QLSRELDTLRGIFRQ	QLSRELDTLRGIFRQ	116	bZIP	330	344	[]	0.91
TLRGIFRQLPESSLV	TLRGIFRQLPESSLV	117	bZIP	337	351	[]	0.93
QLPESSLVKAMGNCA	QLPESSLVKAMGNCA	118	CP	344	358	[]	0.78
QLPESSLVK-me2-AMGNCA	QLPESSLVKAMGNCA	119	CP	344	358	352', 'me2',	0.83
QLPESSLVK-ac-AMGNCA	QLPESSLVKAMGNCA	120	CP	344	358	352', 'ac',	0.50

Table 1: C/EBPα peptides screened for protein interactions with PRISMA. Peptides with a Pearson correlation between replicates < 0.6 in PRISMA experiments were excluded from further analysis and are depicted in grey.

2.14. BioID experiments

Cells were seeded in exponential growth phase at a density of $1x10^6$/ml in media supplemented with 1mM biotin and 1 µg/ml doxycycline (one 15cm cell culture dish per replicate, 4 replicates per experiment). Cells were harvested after 24h by centrifugation and washed twice with ice cold PBS. Cell pellets were resuspended in modified RIPA buffer (lysis buffer: 50 mM Tris-HCl pH 7.2, 150mM, NaCl, 1% NP-40, 1mM EDTA, 1mM EGTA, 0.1% SDS, 1% sodium deoxycholate, freshly added protease inhibitors) and incubated on ice for 20 min. Samples were sonicated with a probe sonicator for 5 pulses and centrifuged at 4°C for 20 min at 20000g. The supernatant was transferred into a fresh tube and an aliquot of the protein extract was saved for protein concentration measurement (protein concentration of samples was 5-6mg/ml) and western blotting. Protein extracts were snap frozen in liquid nitrogen and thawed on ice prior to neutravidin pull-downs. For each pull-down, 80µl neutravidin-agarose bead slurry (Pierce™ NeutrAvidin™ Agarose, Thermo Fisher Scientific) was used. Beads were washed twice with lysis buffer and resuspended again in lysis buffer before being added to the protein extracts. The samples were incubated rotating at 4°C for 2.5h. After incubation, protein extracts were removed and the beads were washed 3x with lysis buffer, 1x with 1M KCl, 1x 2M Urea in 50mM Tris pH 8 and 3x with 50mM Tris pH 8. Washing buffers were kept on ice and each washing step was performed with 1ml, inverting the tube 5 times and then centrifuging for 1 min at 2000g to pellet the beads. The washed neutravidin pull-downs were subjected to on bead digestion.

2.15. On bead digestion

Washed beads were resuspended in 80µl urea/trypsin buffer (2M urea, 50mM Tris pH 7.5, 1mM DTT and 5µg/ml trypsin) and incubated 1h at RT on a thermoshaker at 1000 rpm. The supernatant was transferred into a fresh tube and the beads washed twice with 60µl 2M urea/50mM Tris pH7.5, and the supernatant combined with the previous one. The samples were spun down for 1 min at 5000g to remove residual beads and the supernatant was transferred into a fresh tube. Eluted proteins were reduced with 4mM DTT at RT for 30 min and alkylated with 10mM IAA at RT in the dark for 45 min (both on thermoshaker set to 1000 rpm). For tryptic digests, 0.5µg trypsin was added per sample and samples were incubated overnight (16h) at RT on a thermoshaker set to 700 rpm. For an AspN digest, 0.5µg of trypsin and 0.5µg sequencing grade AspN (Promega) were added to the sample. Following overnight digestion, samples were acidified by adding TFA and desalted with STAGE tips.

2.16. In solution digestion

Proteins were digested into peptides as described before (Dittmar et al., 2019). In brief, proteins were reduced with 1mM TCEP (Sigma-Aldrich, Germany) for 30 min followed by alkylation with 5mM CAA final concentration (Sigma-Aldrich, Germany) for 20 min. Sequencing grade lysyl endopeptidase (lysC), mass spectrometry grade (Fujifilm Wako Chemicals, Japan) was dissolved in MS-grade H2O at 0.5 µg/µl and added to the samples at a ratio of 1:50. Samples were digested for 2h before being diluted with four volumes of 50mM ammonium-bi-carbonate (ABC, Sigma-Aldrich Germany) and addition of sequencing grade modified trypsin (Promega, Germany) at a ratio of 1:50. The digestion was continued overnight at RT (14 h) and digested samples were acidified with 20% trifluoracetic acid (TFA, Sigma-Aldrich, Germany) prior to desalting. SILAC labelled samples that were labelled only with heavy lysine were subjected to digestion with LysC only, following the same protocol without the addition of trypsin.

2.17. Desalting with STAGE tips

Digested peptides were desalted with C18 STop and Go Extraction tips (STAGE tips) as described before (Rappsilber et al., 2003). Briefly, three C18 disks solid phase extraction disks (Empore 3M, USA) were punched out and placed into a 200µl plastic pipette tip. STAGE tips were placed into 2ml eppendorf tubes fitted with a custom-made adaptor. Washing and loading steps of STAGE tips were performed in a benchtop centrifuge at RT for 2 min at 2500g. STAGE tips were wetted with 50µl methanol, followed by washing with 100µl STAGE tips-elution buffer (50% ACN/0.1% FA) and 100µl STAGE tips - washing buffer (2% ACN/0.1% ACN). Samples were loaded followed by three washing steps with 200µl STAGE tips-washing buffer. Peptides were eluted from stage tips with 50µl STAGE tips - elution buffer and lyophilized in a speed vac.

2.18. LC-MS/MS

Desalted and dried peptides were resuspended in MS sample buffer (3% ACN/ 0.1% FA) and separated online with an Easy-nLC™ 1200 coupled to a Q-Exactive+ or a Q-Exactive HF-X mass spectrometer equipped with an orbitrap electrospray ion source (all Thermo Fisher Scientific). Column type, gradient length and MS acquisition method was chosen depending on the type of sample.

PRISMA pull-downs were separated on a 20 cm reverse-phase column (inner diameter 75 µm) packed in house with 3 µm C18-Reprosil beads (Dr. Maisch,

Germany) with a linear gradient ramping from 3% to 76% ACN in 33 min, followed by a plateau at 76% ACN for 5 min and subsequently 60% ACN for 5 min. MS data was acquired on a Q-Exactive+ in data dependent acquisition (DDA) mode with a top10 method. Full scan MS spectra were acquired at a resolution of 70000 in the scan range from 300 to 1700 m/z, automated gain control (AGC) target value of 1e6 and maximum injection time (IT) of 120 ms. MS/MS spectra were acquired at a resolution of 17500, AGC target of 1e5 and maximum IT of 60 ms. Ions were isolated with a 2 m/z isolation window and normalised collision energy (NCE) was set to 26. Unassigned charge states and single charged precursors were excluded from fragmentation and dynamic exclusion was set to 20 s.

Whole proteome samples of SILAC labeled cells were analysed with a 0.1x 200 mm MonoCap C18 HighResolution Ultra column (GL Sciences, Netherlands) with a linear gradient ramping from 3% to 48% ACN in 202 min, followed by a plateau at 76% ACN for 10 min and subsequently 3% ACN for 40 min. MS data was acquired on a Q-Exactive+ in DDA with a top10 method. Full scan MS spectra were acquired at a resolution of 70000 in the scan range from 300 to 1700 m/z, AGC target was set to 3e6 and maximum IT to 20 ms. MS/MS spectra were acquired at a resolution of 17500, AGC target of 1e6 and maximum IT of 60 ms. Ions were isolated with a 2 m/z isolation window and NCE was set to 26. Unassigned charge states and single charged precursors were excluded from fragmentation and dynamic exclusion was set to 30s.

BioID pull-downs were separated on a 20cm reverse-phase column packed in house with 3 µm C18-Reprosil beads (inner diameter 75µm) with a gradient ramping from 2% to 54% ACN in 98 min, followed by a plateau at 72% ACN for 5 min and a subsequent plateau at 45% ACN for 5 min. MS data was acquired on a Q-Exactive HF-X in DDA with a top20 method. Full scan MS spectra were acquired at a resolution of 60000 in the scan range from 350 to 1700 m/z, AGC target was set to 3e6 and maximum IT to 10 ms. MS/MS spectra were acquired at a resolution of 30000, AGC target of 1e5 and maximum IT of 86 ms. Ions were isolated with a 1.6 m/z isolation window and NCE was set to 26. Unassigned charge states and ions with a charge state of one, seven or higher were excluded from fragmentation and dynamic exclusion was set to 30s.

2.19. Targeted MS analysis of R142 methylation

Synthetic heavy peptides with the sequence DGRLEPLEYER and DGRmeLEPLEYER were custom synthesised by JPT (spiketides L, labeled at the C-terminus with heavy arginine (Arg10)). Synthetic peptides were dissolved in 50% ACN/ 50mM ABC (stock solution). Digested and desalted C/EBPα BioID pull-downs

(combined AspN/trypsin digest) were resuspended in MS sample buffer containing 100 fmol/µl of the synthetic peptides and measured on a Q-Exactive HF-X mass spectrometer coupled to an Easy-nLC™ 1200 HPLC system. Peptides were separated on a 60 min gradient ramping from 2% to 76% ACN. MS data acquisition cycled between a Top1 MS/data dependent MS2 and data independent measurement of an unscheduled inclusion list (table 2). Resolution of the PRM scan was 60000 with an AGC target of 1e6, 200 ms maximum IT, 0.7 m/z isolation window and a NCE of 27. PRM data was analysed with the skyline software (MacLean et al., 2010).

Mass [m/z]	Formula [M]	Species	CS [z]	Polarity
424.55	DG[Rme]LEPLYER	heavy	3	Positive
419.88	DGRLEPLYER	heavy	3	Positive
421.22	DG[Rme]LEPLYER	light	3	Positive
416.55	DGRLEPLYER	light	3	Positive

Table 2: PRM m/z inclusion list to detect an R142 methylated CEBPA peptide.

2.20. MS raw data processing with MaxQuant

Mass spectrometry raw files were processed with MaxQuant (Cox and Mann, 2008) (version 1.5.2.8) searching against a human protein database containing manually curated isoforms and further unreviewed entries downloaded from Uniprot (June 2017) and a database including common contaminants. Fixed modifications were set to carbamidomethylation of cysteines and variable modifications set to methionine oxidation and N-terminal acetylation. For BioID experiments, lysine biotinylation was added as an additional variable modification. Depending on digestion mode (trypsin or LysC only), enzyme specificity was selected with a maximum of 2 missed cleavages per peptide. The initial allowed mass deviation of the precursor ion was up to 6 ppm and 20 ppm for fragments. False-discovery rate (FDR) was set to 1% on protein and peptide level. For SILAC measurements the requantify option was enabled and minimum ratio count was set to 2. The label free quantification algorithm (LFQ) built into MaxQuant (Cox et al., 2014) was employed for analysis of BioID and PRISMA data. For LFQ analysis, the match between run and fast LFQ option was enabled and minimum ratio count was set to 2.

2.21. Data analysis of mass spec data

Statistical analysis of the dataset was performed using the R-statistical software package (version 3.4.1). The protein groups output file from MaxQuant was filtered for contaminants, reverse hits and proteins only identified by site.

2.21.1 PRISMA data

PRISMA data was filtered for proteins that were detected at least twice per sample group with at least 2 peptides. Samples with a Pearson correlation of LFQ values between technical replicates < 0.6 were excluded from further analysis (Table 1). Missing LFQ values were imputed from a distribution at the detection limit of the mass spectrometer as described before (Keilhauer et al., 2014). For this purpose, a shifted normal distribution was created for each run. The mean of the shifted distribution was 1.8 standard deviations away from the observed mean and the standard deviation of 0.3 times the observed standard deviation. LFQ data was averaged across all three SILAC channels and analysed with a two sample moderated t-test (Limma package). A specific control group was created for each peptide that contained all other peptides except peptides with a sequence homology > 50%. Resulting p-values were adjusted for each PRISMA sample by multiple testing correction with the Benjamini-Hochberg procedure. The significance cut-offs employed were < 10%FDR and $\log_2$(ratio peptide/control) > 1. The LFQ intensities of significant proteins were normalised across all PRISMA peptides by dividing by the maximum LFQ value of the respective protein (normalisation between 0 and 1 across rows). In order to identify PTM dependent binding, a fold change cut-off was employed: ratios between modified and unmodified peptides were calculated for each replicate, an interactor was considered PTM dependent if the ratio in both replicates was bigger or smaller than 2-fold. PRISMA data was integrated with data from BioID experiments and public C/EBPα interactome data (BioGRID (Chatr-Aryamontri et al., 2017), STRING (Szklarczyk et al., 2015), Giambruno et al., 2013). Interactors were loaded into Cytoscape (Shannon et al., 2003) (V3.7.1) and experimentally validated interactions were retrieved by the STRING plug-in.

2.21.2 BioID data

BioID data was filtered for proteins that were detected at least three times per sample group with at least 3 peptides**.** LFQ values of CEBPA BioID pull-downs were compared against controls (BioID and no Dox control) with a moderated a two sample

moderated t-test (Limma package). The significance cut-offs employed were enrichment against controls with FDR < 5% and $\log_2$(fold change) > 1. For comparison of CEBPA isoform-specific pull-downs (P42 vs P30) and WT vs mutant CEBPA pull-downs the cutoffs were FDR < 10% and enrichment against controls as described above.

2.22. Gene ontology (GO) term and analysis

GO term analysis of significant interactors was performed with the DAVID online functional annotation tool using the default parameters (version 6.8) (Huang et al., 2009). Selected significant GO terms were visualized with as a heatmap with R.

2.23. Single sample gene set enrichment analysis (ssGSEA)

Single sample gene set enrichment analysis (ssGSEA 2.0) (Subramanian et al., 2005) was performed in R with a script retrieved from github (github.com/broadinstitute/ssGSEA2.0). Ranked changes of RNA expression were analysed with immunologic and transcription factor target databases downloaded from the MSig database (Liberzon et al., 2011). In Figure 6 and 23 the normalised enrichment score (NES) of significant (< 5% FDR) positively or negatively enriched gene sets is displayed.

3. Results

The results section of the present thesis consists of three parts: First, myeloid NB4 cells were established as a model system for C/EBPα PPI studies. Second, novel PTM sites on C/EBPα were detected. The third and main part describes the results from PRISMA and BioID experiments as an attempt to delineate the SLiM and PTM dependent interactome of C/EBPα in NB4 cells.

3.1. Establishment of the myeloid NB4 cell line as a model system

PPIs may be cell- and/or context-specific. Given the pivotal role of C/EBPα in myeloid differentiation, the promyelocytic human cell line NB4 was chosen as a model system for studying C/EBPα PPI networks. Since cytosolic proteins are not expected to interact with C/EBPα, nuclear extract was used for the PRISMA screen described here. Nuclear extract preparation from NB4 cells was optimised. NB4 cells can be induced to differentiate into granulocytes or monocytes/macrophages with 12-O-tetradecanoylphorbol-13-acetate (TPA) or all-trans retinoic acid (ATRA) respectively (Lanotte et al., 1991). Including chemically induced or differentiated NB4 in C/EBPα interactome studies may facilitate the detection of PPIs that only occur in this specific context. To make an informed decision on which differentiation time point should be included in subsequent PPI analyses, differentiation of NB4 cells was established and kinetic changes of the proteome, as well as the transcriptome, were monitored. In addition, the impact of C/EBPα-BioID expression in NB4 cells on gene expression and differentiation was evaluated.

3.1.1. Differentiation of NB4 cells can be induced with ATRA and TPA

NB4 cells were treated with 50nM TPA or 2μM ATRA and successful differentiation was monitored by analysis of myeloid surface marker and morphological staining (**Figure 5**). The myeloid surface marker CD11b (or integrin alpha M (ITGAM)) is expressed on granulocytes, monocytes/macrophages, and natural killer cells but not on myeloid precursor cells. CD11b surface expression was highest after two days of TPA or 6 days of ATRA treatment respectively, with the fraction of CD11b positive cells reaching over 90% (**Figure 5A**). TPA treated cells started to become adherent within hours of treatment and cytofluorimetric analysis of forward and side scatter also revealed increased granularity and size of TPA treated cells and slightly increased granularity of ATRA treated cells (**Figure 5B**). Histological Wright Giemsa staining confirmed expected morphological changes: increased size and cytoplasm of NB4

derived monocytes/macrophages and typical segmentation of nuclei in NB4 derived granulocytes (**Figure 5C**).

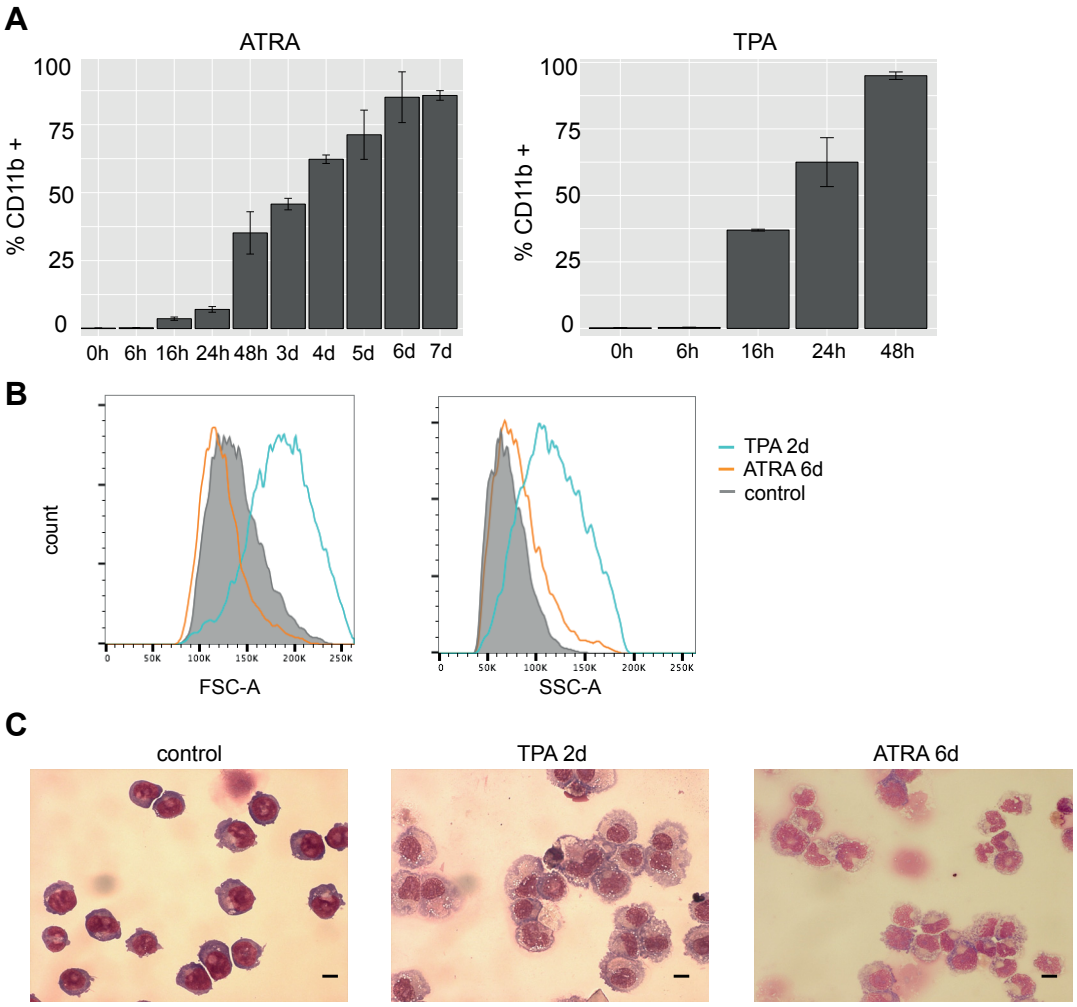


Figure 5: ATRA and TPA induce morphological and surface marker changes in NB4 cells.
NB4 cells were treated with 2µM ATRA or 50nM TPA and analysed at specified time-points. **A**: CD11b surface marker expression measured by FACS analysis
B: TPA and ATRA treatment induce changes in cell size (forward scatter (FSC)) and granularity (side scatter (SSC)). **C**: Histological Wright Giemsa staining. Scale bar represents 10 µM.

3.1.2. Proteomic and transcriptomic changes during NB4 differentiation

NB4 cells were treated with ATRA/TPA as indicated in **Figure 6A** and RNA and proteins were extracted at different time points. Kinetic changes in the proteome and transcriptome of NB4 cells undergoing differentiation were monitored with SILAC based mass spectrometric analysis (label swap experiment) and microarray RNA profiling (biological triplicates). Biological replicates clustered together in a principal component analysis (PCA) that separates proteomic and transcriptomic data in time (component 1) and treatment (component 2) (**supplemental Figure 1**). During differentiation, vast transcriptomic and proteomic changes occur in NB4 cells and around a third of all detected transcripts and proteins significantly change in at least one of the time points analysed (**Figure 6B**). Gene set enrichment analysis (GSEA) of transcriptomic and proteomic changes revealed the regulation of several pathways associated with myeloid maturation (**Figure 6C**). As cells differentiate, cell growth and correspondingly also DNA templated transcription are down-regulated while cell death and immune system related processes are up-regulated. During myeloid differentiation, cells undergo morphological changes as reflected by the up-regulation of cytoskeleton organisation related genes. Autophagy is essential to several functions of mature myeloid cells including reactive oxygen species and cytokine production as well as degranulation. ATRA and TPA treatment also induce changes in several signalling cascades like the Wnt and the Erbb signalling pathways as well as PI3K activity.

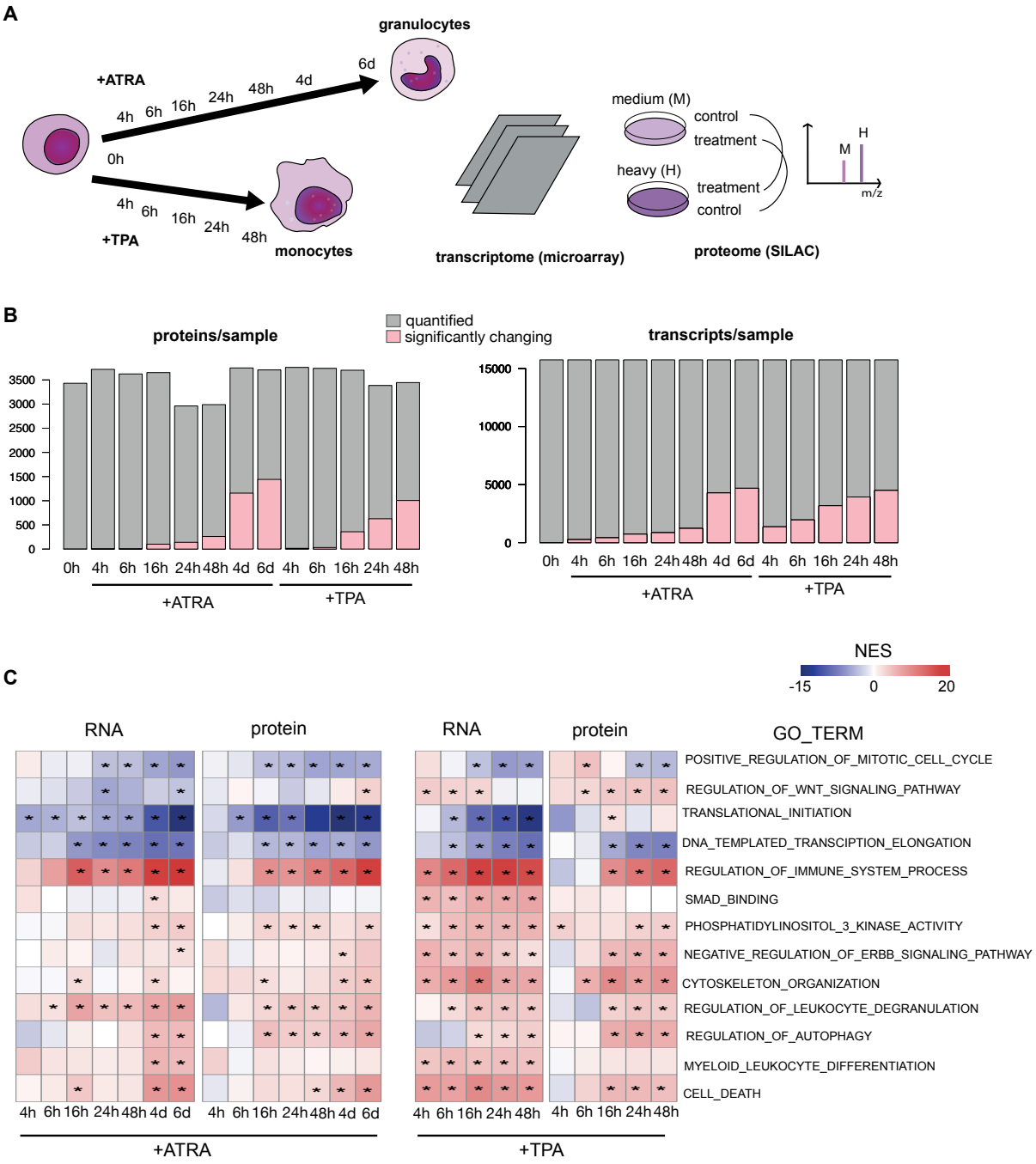
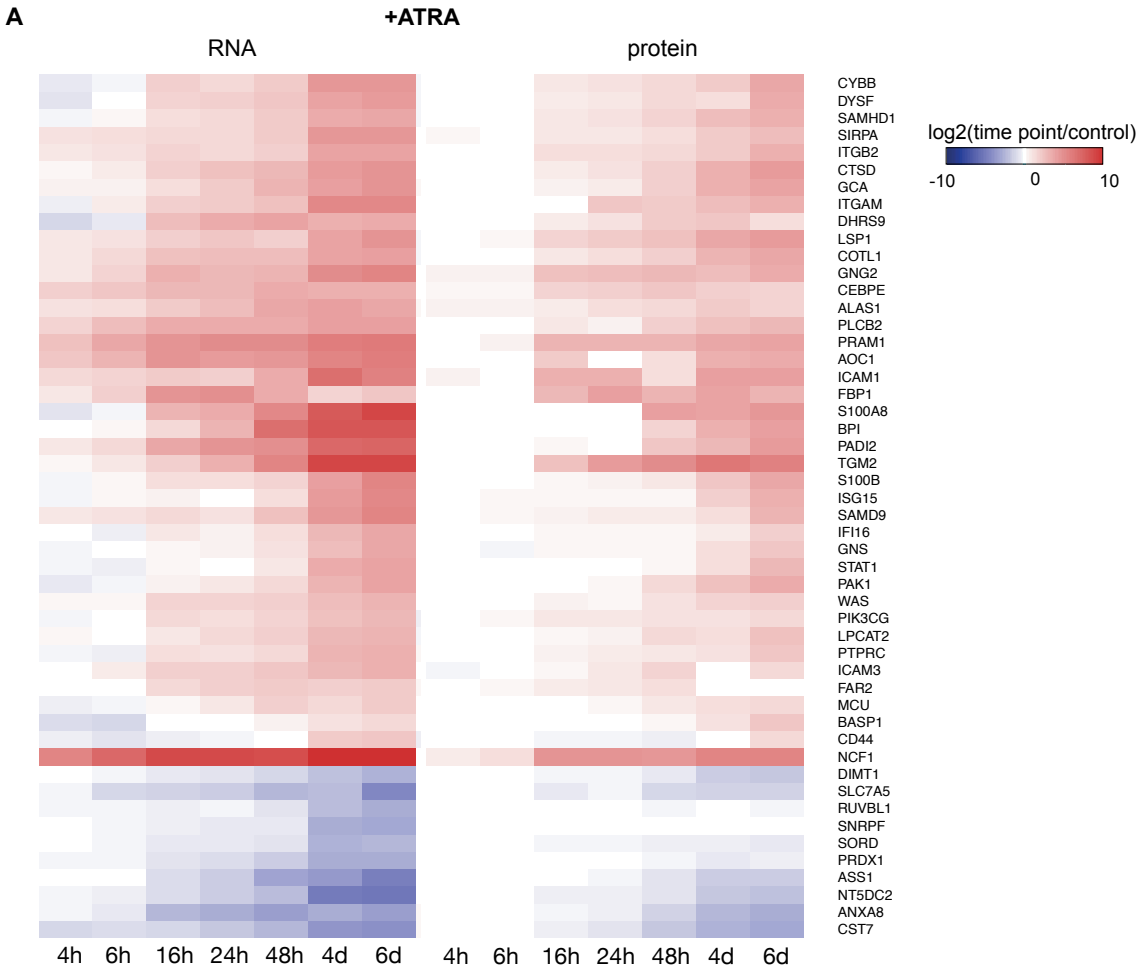


Figure 6: Kinetic proteome and transcriptome changes during NB4 differentiation. **A:** NB4 cells were treated with 2µM ATRA or 50nM TPA and analysed at the indicated time points with SILAC based proteomics (n=2) and microarray RNA expression analysis (n=3). **B**: Number of quantified and significantly changing features in each sample (significance cut-offs proteome was $\log_2$(fold change) in both replicates >1 or < -1; significance cut-offs transcriptome was < 1%FDR) . **C**: GSEA analysis of proteomic and transcriptomic changes relative to control. Color code represents the normalised enrichment score (NES). Significant GO terms (< 5% FDR) are indicated with an asterisk.

The 50 most differentially regulated transcripts and their corresponding proteins are displayed as a heatmap in **Figure 7A,B**. Among up-regulated genes in TPA treated cells are several integrins for cell-extracellular matrix adhesion and matrix metalloproteinases (MMPs) that play an important role in tissue remodelling. The down-regulation of myeloperoxidase (MPO) upon TPA treatment is in concordance with findings from other myeloid cells (Zheng et al., 2002). Positive response to ATRA treatment and granulocytic differentiation is reflected by the up-regulation of neutrophil cytosolic factor 1 (NCF1), PML-RARA-regulated adapter molecule 1 (PRAM1) and the granulocytic transcription factor C/EBPε. Changes of C/EBPα protein and RNA expression in NB4 differentiation are displayed in **Figure 7C**. C/EBPα protein levels peak after 6h of ATRA treatment and are down-regulated in the later time points. In TPA treated NB4 cells C/EBPα is immediately down-regulated and goes up again at 48h. The C/EBPα peptides that were detected here are not specific for C/EBPα isoforms and therefore changes on C/EBPα isoform abundance cannot be described. Based on the results, NB4 cells treated for 6h with TPA or 12h ATRA, as well as control cells, were included in the PRISMA C/EBPα screen described in the following sections



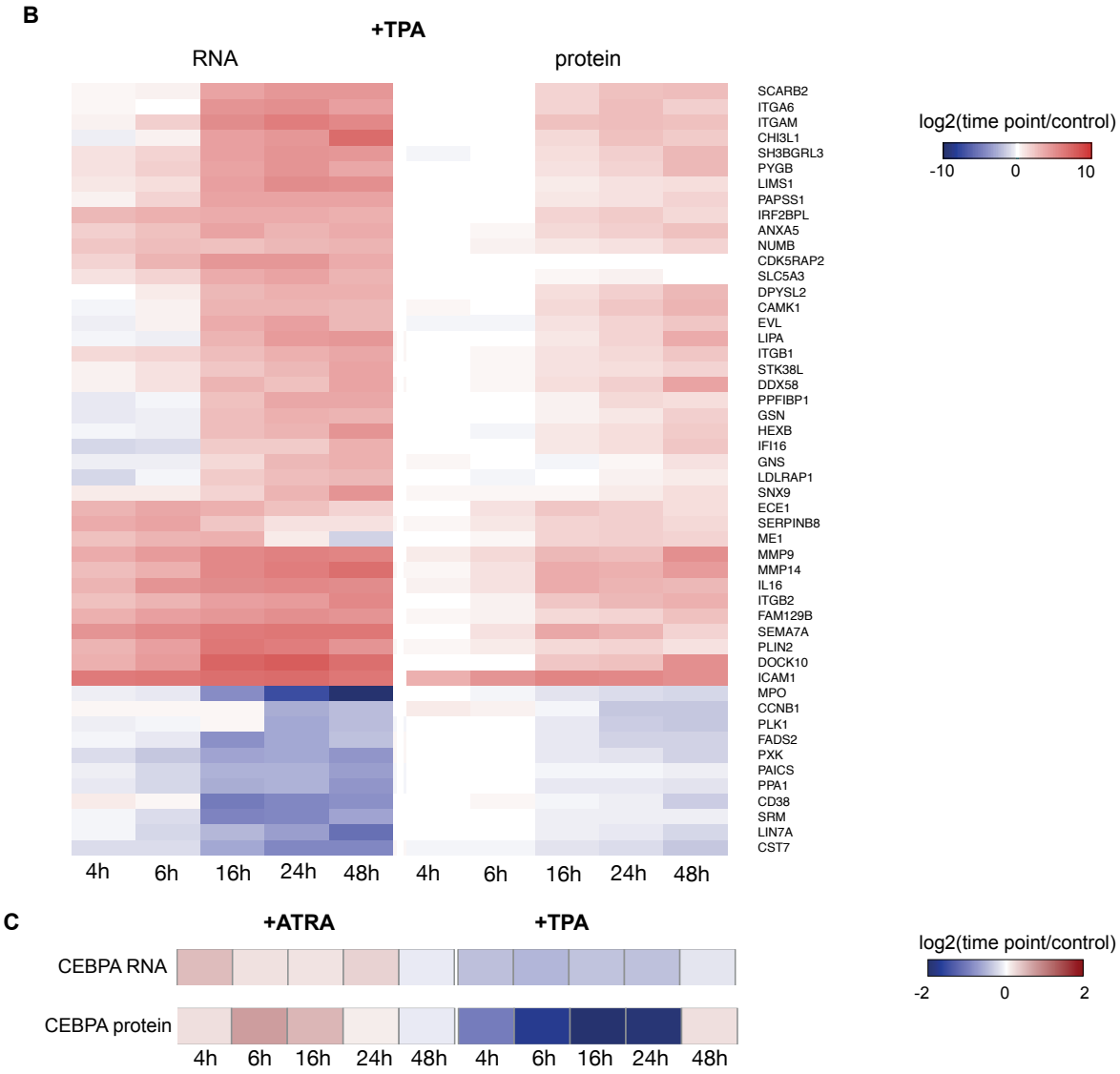
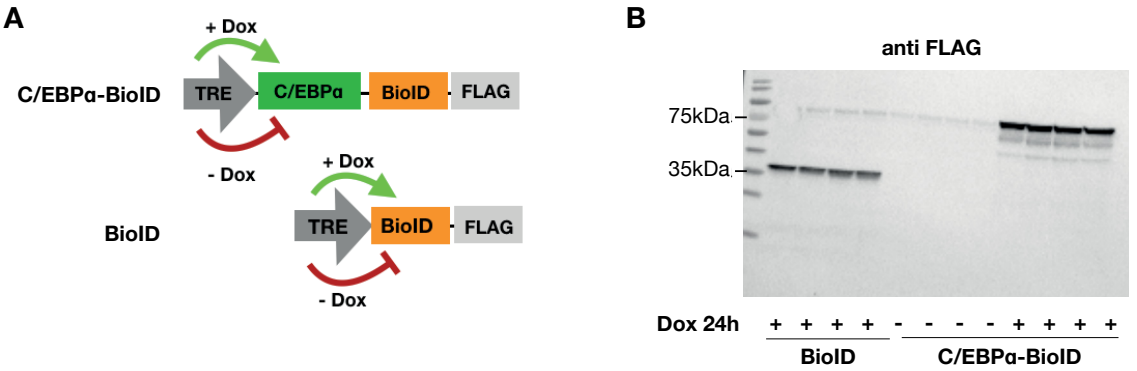


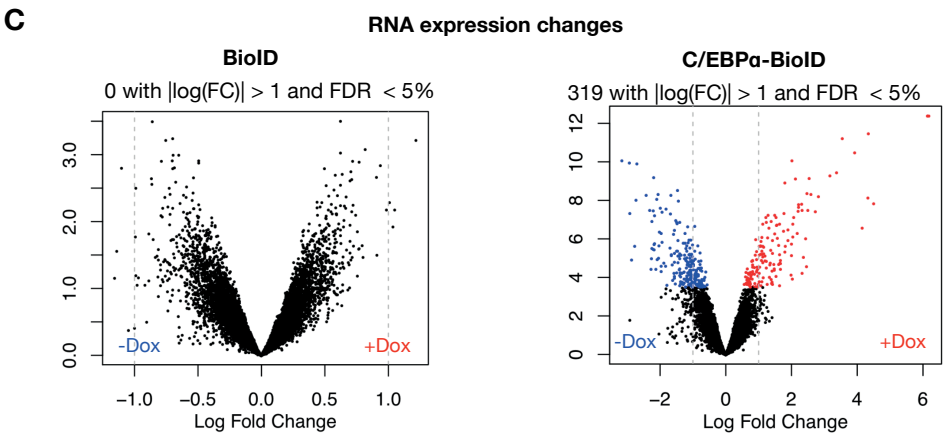
Figure 7: Transcripts and proteins regulated during NB4 differentiation.
Heatmap represents log_2(fold changes) relative to control. **A:** Most regulated transcripts (by FDR) and corresponding proteins in ATRA induced differentiation.
B: Most regulated transcripts (by FDR) and corresponding proteins in TPA induced differentiation. **C**: RNA and protein expression changes of C/EBPα during NB4 differentiation.

3.1.3. Expression of C/EBPα-BioID in NB4 cells induced target genes but did not induce terminal differentiation

NB4 cells were genetically engineered to express BioID or C/EBPα C-terminally fused to BioID (C/EBPα–BioID) under the control of a doxycycline inducible promoter (**Figure 8A**). An IRES GFP in the lentiviral vector enabled FACS sorting of successfully transduced cells. Induction of C/EBPα–BioID and BioID was confirmed via western blotting and detection with an antibody directed against the C-terminal FLAG-tag (**Figure 8B**). RNA from doxycycline induced and control cells was extracted and subjected to microarray analysis (n=3). In total, 391 genes were regulated by C/EBPα–BioID expression (FDR < 1%, |FC| < 0.5) while BioID expression alone did not induce any changes in gene expression as expected (**Figure 8C**). In C/EBPα–BioID expressing NB4, most significantly up-regulated genes included defensins (DEFA3, DEF1A, DEF1B), S100A9 and S100A8, which are all known C/EBPα targets (Birkenmeier et al., 1989). This indicates that C/EBPα–BioID is functional as a transcription factor in NB4 cells. Although ectopic C/EBPα has been described to induce differentiation of several cell lines (Huafeng et al., 2004; Porse et al., 2001; Radomska et al., 1998), expression of C/EBPα–BioID in NB4 did not increase the fraction of CD11b positive cells (**Figure 8D**). The response of NB4 cells to ATRA treatment was not affected by C/EBPα–BioID expression.

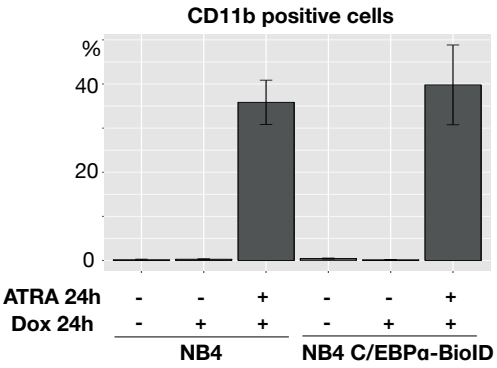
Figure 8: Stable NB4 cell lines inducibly express C/EBPα-BioID or BioID.
A: Stable NB4 cell lines were engineered to express C/EBPα-BioID or BioID under the control of a tet-responsive element (TRE) **B:** Successful induction was confirmed by western blotting. **C**: RNA expression changes induced by expression of C/EBPα-BioID or BioID in NB4 cells. Cells were treated for 24h with or without doxycycline (Dox) and RNA expression was analysed by microarray (n=3). Table indicates the top25 regulated genes by C/EBPα-BioID expression. No genes were regulated by BioID expression. **D:** CD11b surface marker expression of NB4 cell lines treated as indicated (n=3).





C/EBPα-BioID top25 regulated genes

GeneSymbol	logFC	adj.P.Val
DEFA3	6.19	1.71E-09
DEFA1B;DEFA1;DEFA3	6.15	1.71E-09
DEFA1;DEFA1B	6.15	1.71E-09
SERPINB2	4.34	1.08E-08
S100A9	3.92	6.98E-08
ENTPD1	3.55	1.48E-08
TNFAIP6	4.51	6.32E-06
CDA	4.33	3.88E-06
RAB7B	-3.16	1.34E-07
S100A8	3.38	4.05E-07
CD3D	-2.93	1.55E-07
MS4A7	3.17	5.32E-07
GJA1	-2.70	1.55E-07
ANKRD22	4.15	5.20E-05
MCU	2.54	6.13E-07
CARD16	2.83	3.41E-06
CD300A	-2.72	4.46E-06
LYZ	2.60	3.04E-06
LGALS12	-2.92	1.37E-05
GRAMD1B	2.02	1.34E-07
ANXA1	2.48	2.94E-06
LPXN	-2.19	6.10E-07
PCOLCE2	2.74	1.19E-05
GATA2	-2.41	3.09E-06
CD38	2.13	6.13E-07



3.1.4. Nuclear extract preparation from NB4 cells

Nuclear extracts from NB4 cells were prepared according to a protocol adapted from Dignam et al., 1983. Since the commonly used detergent NP-40 is not compatible with subsequent mass spectrometry applications, it was replaced with the more compatible detergent n-Dodecyl β-D-maltoside (DDM) (Laganowsky et al., 2014). Successful enrichment of nuclear proteins was confirmed by western blotting and detection of nuclear and cytosolic loading controls (histone H3 and actin, **Figure 9**). The obtained protein concentration of nuclear extract was between 5 and 6 mg/ml.

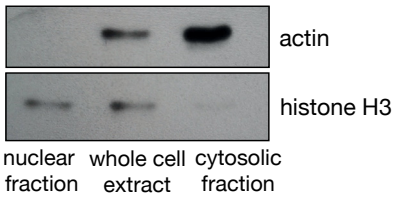


Figure 9: Nuclear extract preparation from NB4 cells.
Purity of the nuclear fraction was confirmed by western blotting and detection of nuclear and cytosolic loading controls.

3.2. Identification C/EBPα PTM sites

Arginine methylation of both C/EBPα and C/EBPβ is implicated in the regulation of protein interactions (Leutz et al., 2011). The CR1L region of C/EBPα contains three evolutionary conserved arginine residues (R142, R147, R154) that are potential methylation targets. Site-directed mutagenesis of these arginine residues altered the transdifferentiation potential of C/EBPα by a mechanism possibly connected to PPIs (unpublished data). However, of the three arginines in C/EBPα CR1L, only the mono and dimethylation of R154 has been previously confirmed by an MS/MS spectrum (Li-ming et al., 2019).

Especially the detection of R142 methylation by mass spectrometry is technically challenging – the tryptic peptide spanning the methylation would be over 30 amino acids long and undetectable with traditional LC-MS/MS setups. Therefore, an alternative approach that combined using an alternative protease and targeted mass spectrometry was employed (**Figure 10**). A C/EBPα-BioID pull-down was digested with AspN and a methylated and the corresponding unmodified peptide spanning R142 were monitored with parallel reaction monitoring (PRM). Identity of the unique C/EBPα peptide with the sequence DGRmeLEPLYER was confirmed with a heavy peptide standard, that elutes at the same time of the chromatogram and displays the same fragmentation pattern. The non-methylated peptide spanning R142 (DGRLEPLYER) elutes a few minutes earlier, which is expected since methylation increases

hydrophobicity and affinity of peptides to the C18 material. The measured intensity of the non-methylated peptide was around three orders of magnitudes higher than the intensity of its methylated counterpart. Although peptide intensity does not perfectly correlate with abundance, this indicates that roughly 0.1% of the peptide was methylated in the sample.

On top of performing targeted measurements, untargeted measurements of C/EBPα-BioID pull-downs were analysed including several variable modifications (phosphorylation, acetylation, methylation) as search parameters. In addition to detecting known C/EBPα PTM sites, the arginine residue R12 within C/EBPα CR2 was identified as novel methylation and dimethylation site (**Figure 11**). Except for the mass to charge ratio of peptide fragments that contain the modified R12 residue, the fragmentation pattern of non-methylated, methylated and dimethylated peptide are identical. Together with another 38 PTMs, summarised in **table 3**, R12 and R142 methylation were incorporated in a C/EBPα PRISMA screen to detect their influence on protein interactions.

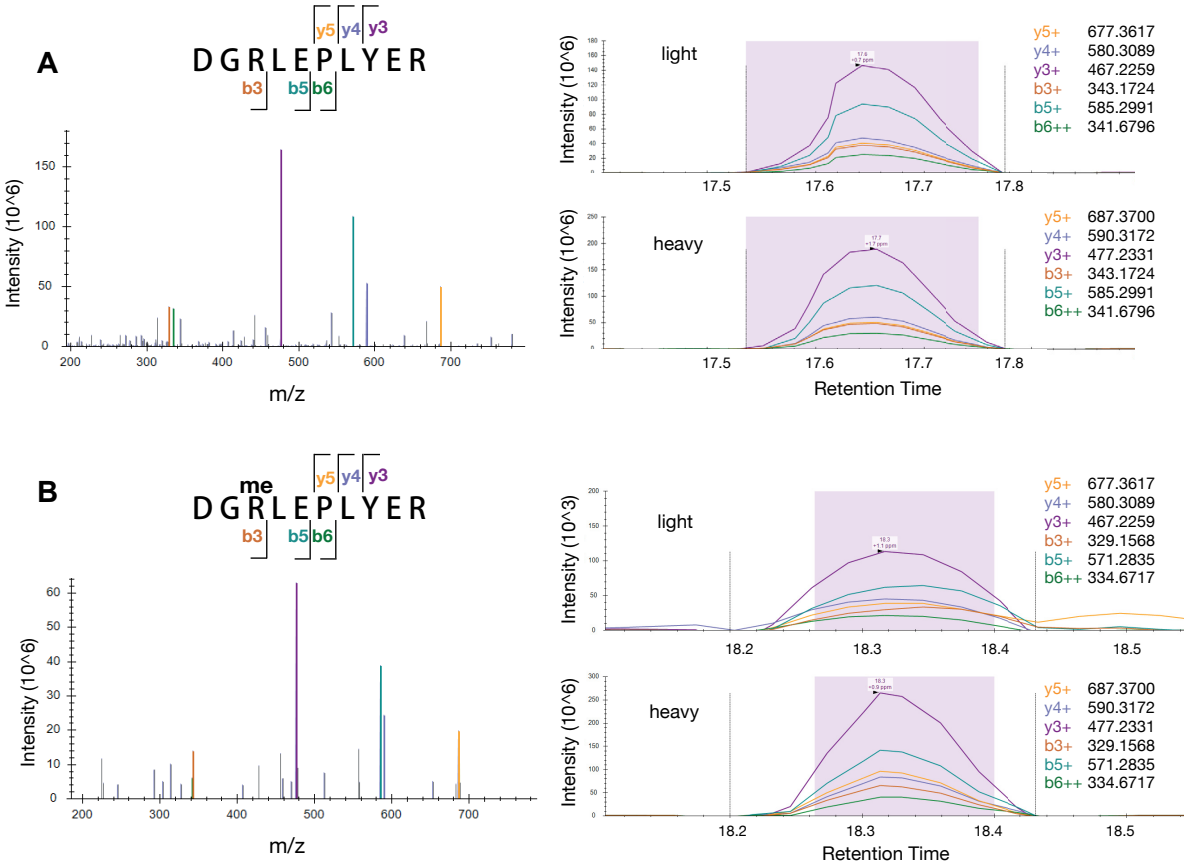


Figure 10: PRM measurements confirmed C/EBPα methylation at R142.
C/EBPα-BioID pull-downs were digested with trypsin and AspN and subjected to PRM measurements specifically monitoring a C/EBPα peptide spanning R142. Identity of the R142 methylated and the R142 unmodified peptide was confirmed with a heavy synthetic peptide standard (labeled at the C-terminus with Arg10). Peptide fragments and their m/z monitored are indicated. **A**: unmodified peptide. **B**: peptide methylated at R142

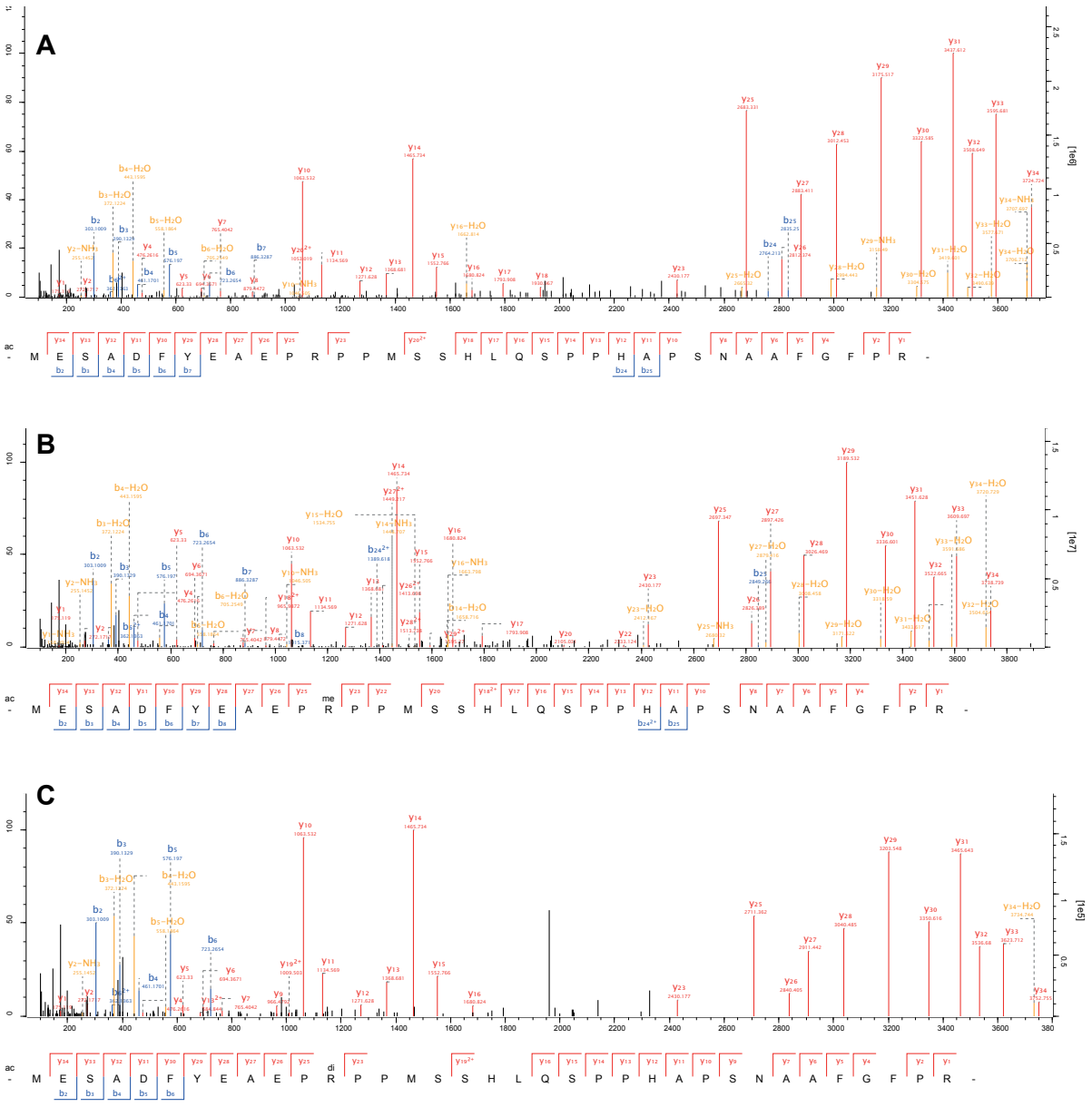


Figure 11: C/EBPα is methylated and dimethylated at R12.
C/EBPα-BioID pull-downs were digested with trypsin and AspN and subjected to shotgun mass spectrometry. MS/MS spectra of an N-terminal C/EBPα peptide spanning R12 are displayed. **A**: no R modification **B**: R12 methylation **C**: R12 dimethylation

siteID	Position(Human) CEBPA P42	Position(rat) CEBPA P42	PTM included PRISMA screen	Known PTM/Reference (Human)	Known PTM/Reference (Rat)	known PTM features	BioID detected	phenotype
1	R12	R12	me2sym, me2asym				me, me2	
2	S21	S21	phosphorylation	PMID:20101026, 16446383, 14701740	PMID:26515730, 16807249	phos	phos	disrupts granulocytic differentiation
3	R35	R35	me2sym, me2asym	PMID:31015230		me	me	increased cyclin D1 expression and proliferation
4	K86	K86	ac,me2					
5	K90	K90	ac,me2					
6	K92	K92	ac,me2					
7	R142	R140	me2sym, me2asym				me	
8	R149	R147	me2sym, me2asym					
9	R156	R154	me2sym, me2asym	PMID:31015230		me, me2	me	increased cyclin D1 expression and proliferation
10	K161	K159	ac, me2	PMID:19608861 PMID:21890473	https://doi.org/10.18452/17281 PMID:22902405 PMID:23639777	me ac ub sumo	ac	interaction with HDAC3
11	R165	R163	citr, me2sym, me2asym		https://doi.org/10.18452/17281	citr		
12	K171	K169	ac, me2					
13	S190	S193	phos	PMID:27619993,26 15376, 15107404		phos		growth inhibition
14	T226	T222	phos	PMID:16600022	PMID:26153766 , 27619993 , 16600022	phos		regulation of lipogenic and gluconeogenic gene expression
15	T230	T226	phos	PMID:16600022	PMID:16600022	phos		regulation of lipogenic and gluconeogenic gene expression
16	S234	S230	phos	PMID:20101026	PMID:17290224, 10567568	phos		
17	K254	K250	ac,me2					
18	R264	R260	me2sym, m2asym					
19	S266	G262	phos	PMID:25159151		phos		
20	K352	K352	ac,me2					

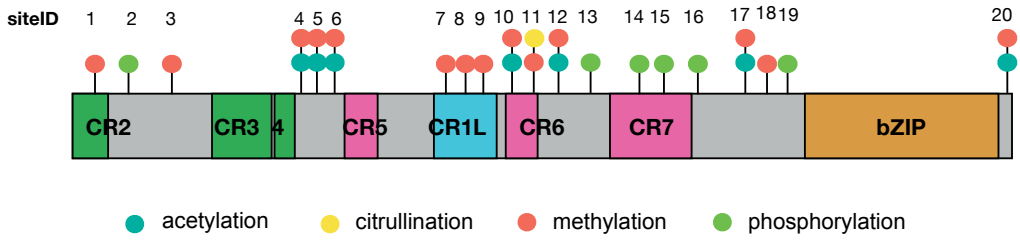


Table 3: C/EBPα modifications included in PRISMA.

3.3. C/EBPα interactome studies

3.3.1. C/EBPα PRISMA screen

In order to explore the linear interactome of C/EBPα and PTM dependencies, the previously published PRISMA method (Dittmar et al., 2019) was slightly modified and adapted for C/EBPα. In brief, the C/EBPα amino acid sequence was split into peptides of 15 amino acid length that were custom synthesised on a cellulose membrane. Peptides were designed with a 7 amino acid sequence overlap (tiling peptides) and included peptides with modifications. In total, 120 peptides were included (maximum of one PTM/peptide) in the screen (**Table 1** in the material and methods section). The C/EBPα peptide matrix was incubated with nuclear extract from SILAC-labeled NB4 cells mixed at equal ratios. Prior to harvesting, heavy labeled cells were treated for 6h with ATRA, medium labeled cells were treated for 6h with TPA and light labeled cells were treated with solvent only. After incubation and washing, individual peptide spots were punched out and bound proteins were prepared for analysis with LC MS/MS. The screen was performed as technical duplicate. A schematic representation of the workflow is depicted in **Figure 12**.

Raw data was processed using the label-free quantification (LFQ) algorithm from MaxQuant (Cox et al., 2014) and LFQ channels were averaged over all three SILAC channels. In total 2274 proteins were identified with an average of 981 proteins detected in each peptide spot. The overall technical reproducibility of the PRISMA screen showed a median Pearson correlation coefficient between replicates of 0.85 and patterns of correlation between different C/EBPα regions (**Figure 13A**). Peptide spots with a Pearson correlation coefficient between replicates < 0.60 (8% of peptides) were excluded from further analysis.

To distinguish specific from unspecific binding events, a moderated t-test was employed (peptide vs. other peptides). The plot in **Figure 13B** shows the FDR of identified proteins in the PRISMA peptide spot ISAYIDPAAFNDEFL plotted over their ratio (peptide/other peptides). In total, 785 proteins passed the significance threshold (< 10% FDR) in at least one peptide spot. LFQ intensities of significant proteins were normalised between 0 and 1 across all PRISMA peptides to reveal binding profiles of interacting proteins across the entire C/EBPα amino acid sequence and PTM sites. The tiling-based filter that was used in a previous PRISMA study (Dittmar et al., 2019) was not applicable for this study, since amino acid shift in the present screen is 7 amino acids as opposed to 4 amino acids in the study by Dittmar et al. Short linear motifs are typically between 4 and 11 amino acids long (Tompa et al., 2014) and a shift of 7 amino acids likely surpasses the motif in most cases. Others have employed a similar

student's t-test based approach for the analysis of large-scale AP-MS datasets (Hein et al., 2015; Keilhauer et al., 2014). Similarly, Meyer et al. have used a rank-based test to compare peptide pull-downs amongst each other and distinguish specific from unspecific binding (Meyer et al., 2018).

In the PRISMA screen described here, the highest number of protein interactions was found within CR2, CR3, CR4 and CR1L that correspond to the major transactivating regions of C/EBPα (**Figure 13D)**. Global binding profiles are visualised as heatmap in **Figure 13C** and sums of normalised intensities are depicted as barplot on top. Extracted binding profiles in **Figure 14** display the interaction profiles of selected proteins and complexes that were significantly enriched in PRISMA and have previously been shown to interact with C/EBPα

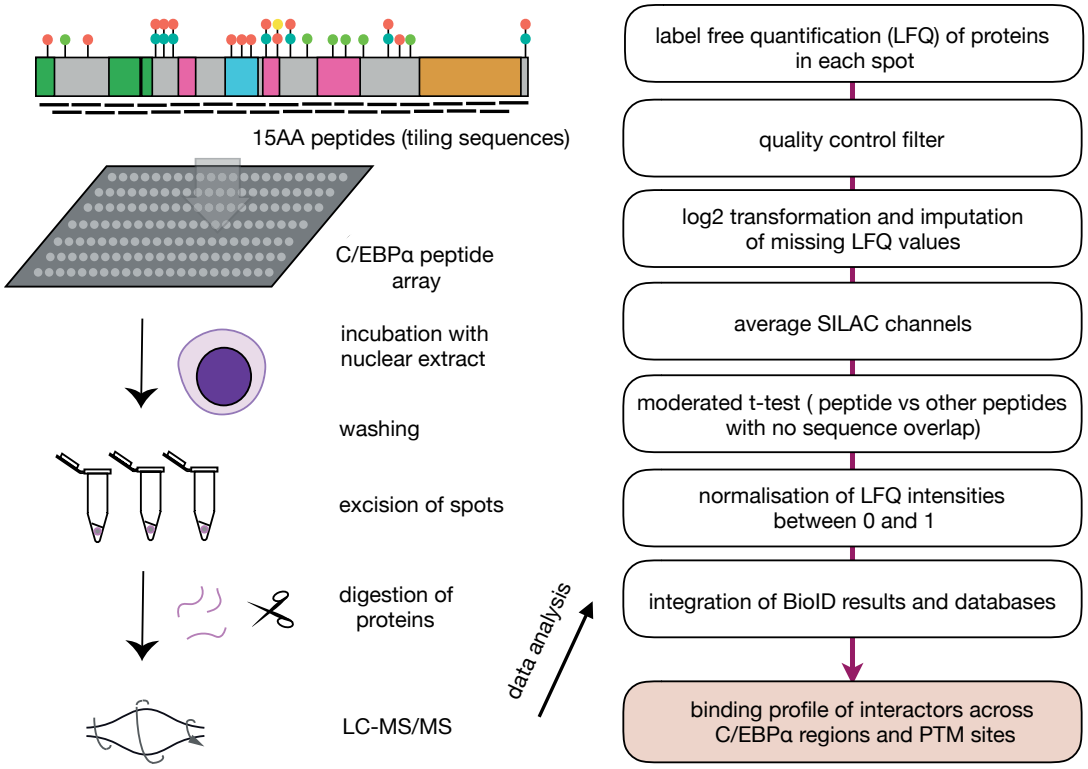
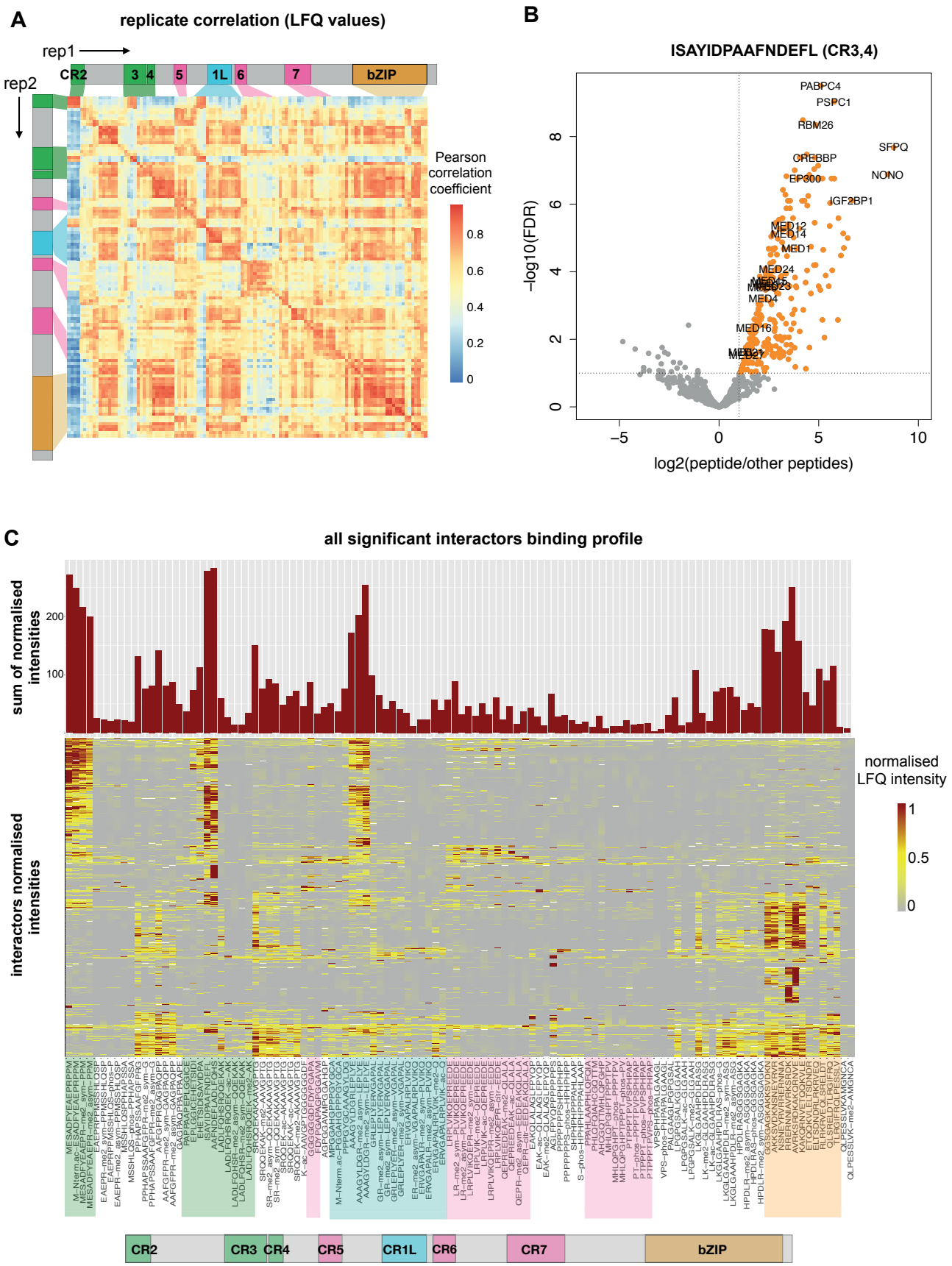


Figure 12: Schematic representation of PRISMA workflow.
A cellulose membrane containing over 120 C/EBPα derived 15 AA long peptides was incubated with nuclear extract from SILAC labeled NB4 cells. After incubation and washing steps the individual peptide spots were punched out and placed separately into a 96 well plate. Bound proteins in each peptide spot were digested and subjected to analysis with LC MS/MS. The screen was performed as a technical duplicate, the data analysis strategy is depicted on the right and described in detail in the materials and methods section.


A
replicate correlation (LFQ values)
rep1
rep2
CR2
3
4
5
1L
6
7
bZIP
Pearson correlation coefficient
0.8
0.6
0.4
0.2
0
B
ISAYIDPAAFNDEFL (CR3,4)
PABPC4
PSPC1
RBM26
SFPQ
CREBBP
NONO
EP300
IGF2BP1
MED12
MED14
MED1
MED24
MED4
MED16
−log10(FDR)
log2(peptide/other peptides)
C
all significant interactors binding profile
sum of normalised intensities
interactors normalised intensities
normalised LFQ intensity
CR2
CR3
CR4
CR5
CR1L
CR6
CR7
bZIP

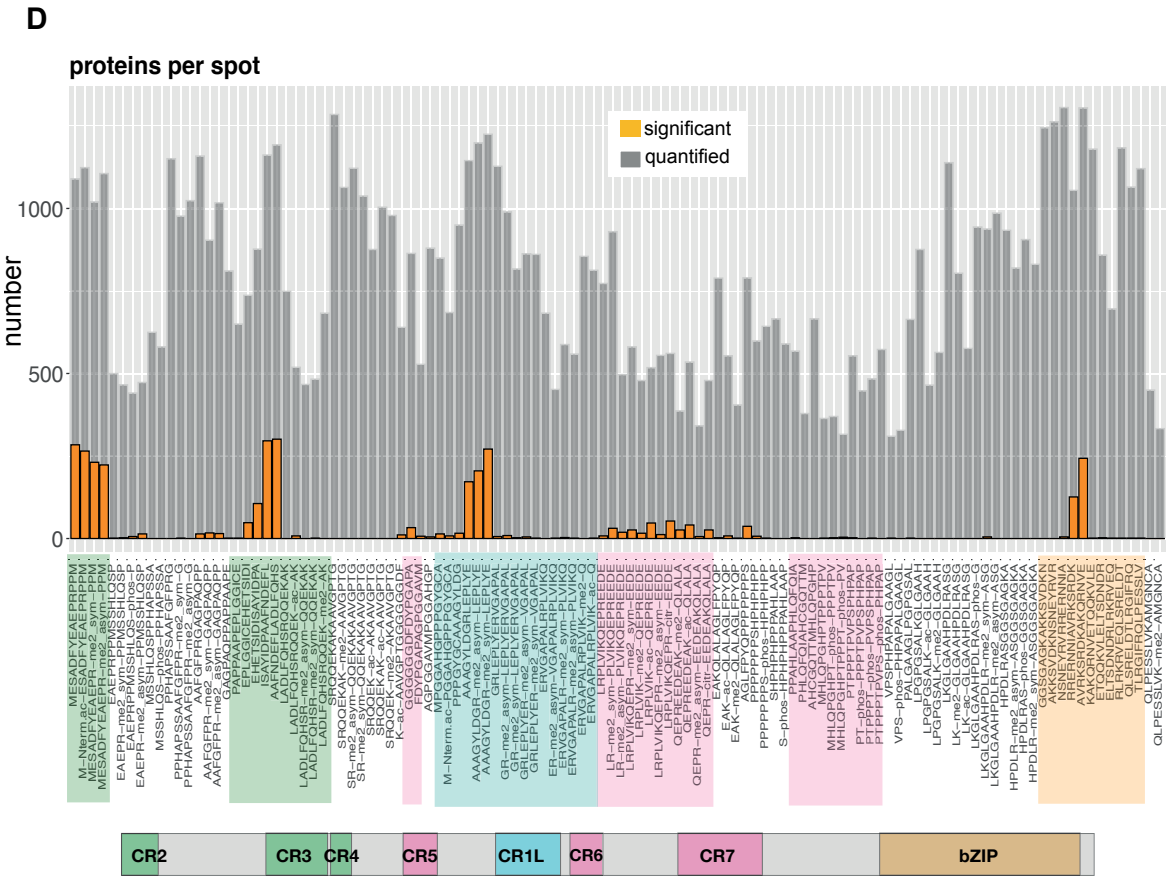
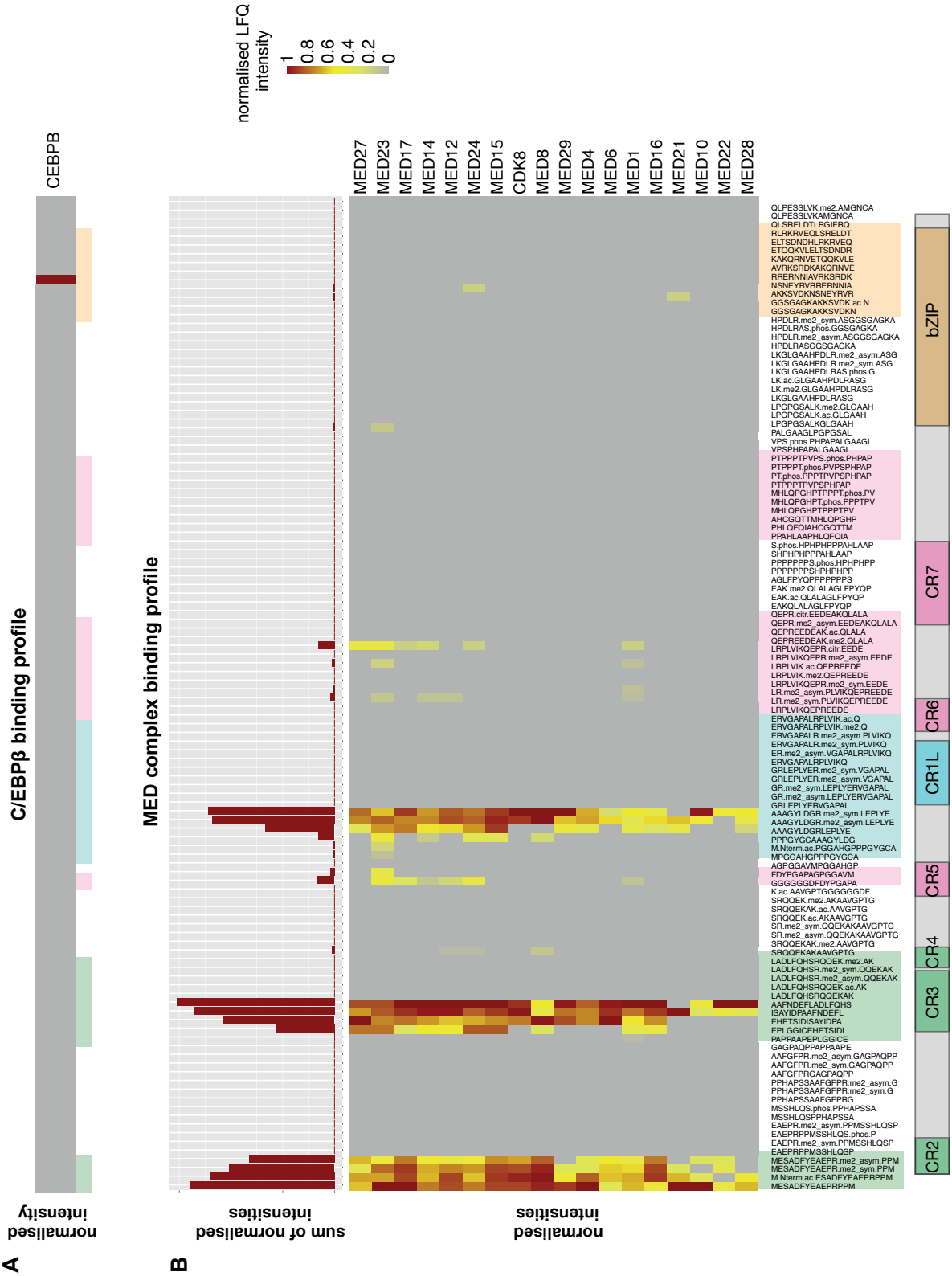


Figure 13: PRISMA facilitated mapping of C/EBPα interactors to the C/EBPα sequence.
A: Pearson correlation matrix of technical replicates. Samples from replicate one (columns) are plotted against samples from replicate two (rows), Samples are ordered from C/EBPα N- to C-terminus. Conserved regions are indicated. **B:** Exemplary analysis of a PRISMA peptide spot with the sequence ISAYIDPAAFNDEFL. Replicates were compared against other peptides with no sequence overlap with a moderated t-test. Enrichment of proteins is plotted against the $-\log_{10}$(FDR). Significance cut-offs and Top8 interactors as well as mediator complex subunits are indicated. **C**: PRISMA binding profile of all proteins (y axis) that pass the significance threshold in at least one PRISMA peptide (x axis). LFQ intensities were normalised between 0 and 1 and are displayed as a heatmap. PRISMA peptides are ordered from C/EBPα N- to C-terminus. Barplot on top of the heatmap represents sums of normalised intensities in each spot.
D: Quantified and significant proteins in each PRISMA peptide spot. PRISMA peptides are ordered from C/EBPα N- to C-terminus.

3.3.1.1. PRISMA binding profiles of known C/EBPα interactors

The mediator of transcription complex **(MED)** is an essential transcriptional coactivator in eukaryotes that interacts with RNA PolII and transcription factors (Soutourina, 2018). All MED proteins displayed similar binding patterns across C/EBPα PRISMA peptides peaking in CR2, CR3/4 and CR1L (**Figure 14B**). The histone acetlytransferases P300-CBP are well described C/EBPα interactors (Erickson et al., 2001; Zaini et al., 2018). In PRISMA, P300-CBP bound most strongly to peptides spanning conserved regions CR3/4 with some residual binding in CR2 and CR1L (**Figure 14D**). This binding pattern is in line with previous findings showing that a region spanning amino acids 55–108 in C/EBPα (corresponding to CR3 and CR4) is sufficient, but not essential to mediate interaction with P300 and to induce adipogenesis (Erickson et al., 2001). Previously the nucleosome remodelling and histone deacetylation (NuRD) complex has been shown to interact with C/EBPβ (Dittmar et al., 2019) and several components (HDAC1/2, MTA2) have also been identified in C/EBPα AP-MS experiments (Grebien et al., 2016). In PRISMA, NuRD subunits interacted with C/EBPα peptides derived from CR3/4 and adjacent regions of low complexity located between CR5 and CR1L and the bZIP domain (**Figure 14C**). Like other C/EBP factors, C/EBPα contains a bZIP domain for dimerisation and DNA binding. In PRISMA, C/EBPβ was significantly binding to a C/EBPα peptide spanning AA 288-302, which corresponds to a region within the bZIP domain **(Figure 14A)**. Misidentification of C/EBPβ due to C/EBPα peptides coming from the PRISMA membrane was excluded by confirming uniqueness of identified C/EBPβ peptides.

Figure 14: PRISMA maps known interactors across the C/EBPα sequence.
PRISMA binding profile of selected proteins and complexes that have previously been shown to interact with C/EBPα. LFQ intensities were normalised across PRISMA peptides between 0 and 1. **A**: C/EBPβ **B:** Mediator of transcription (MED) complex
C: Nucleosome Remodelling Deacetylase (NuRD) complex.
D: Histone acetyltransferases EP300 and CREBBP

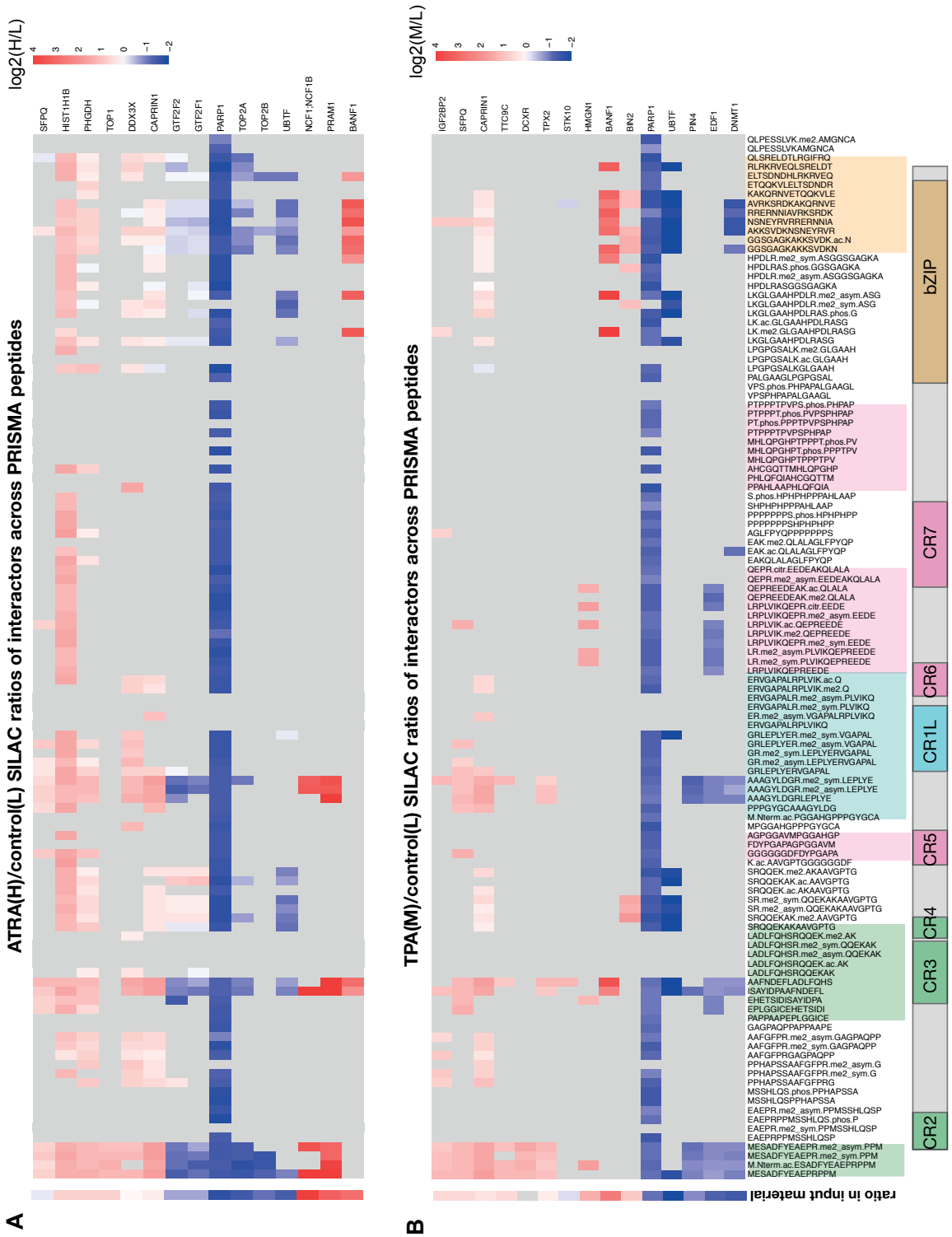

A
C/EBPβ binding profile
normalised intensity
CEBPB
B
MED complex binding profile
sum of normalised intensities
normalised intensities
normalised LFQ intensity
1
0.8
0.6
0.4
0.2
0
MED27
MED23
MED17
MED14
MED12
MED24
MED15
CDK8
MED8
MED29
MED4
MED6
MED1
MED16
MED21
MED10
MED22
MED28
bZIP
CR7
CR6
CR1L
CR5
CR4
CR3
CR2


normalised LFQ intensity
1
0.8
0.6
0.4
0.2
0
C
NuRD complex binding profile
CHD4
HDAC2
GATAD2A
MTA2
HDAC1
normalised intensities
sum of normalised intensities
D
P300-CBP binding profile
EP300
CREBBP
normalised intensities
sum of normalised intensities
bZIP
CR7
CR6
CR1L
CR5
CR4
CR3
CR2

3.3.1.2. Differential PRISMA binding profile of TPA/ATRA/control treated cells

The C/EBPα PRISMA screen described here was performed with SILAC labeled extract from TPA (medium, M) or ATRA (heavy, H) treated as well as control (light, L) NB4 cells. Differential binding patterns of H/M/L labeled proteins were evaluated by investigating SILAC ratios of interacting proteins across PRISMA peptides. SILAC ratios of the 15 most differential PRISMA interactors are displayed as a heatmap across PRISMA peptides in **Figure 15**. Comparison with SILAC ratios from the input material (first column in Figure 15) reveals that high or low SILAC ratios in the PRISMA peptide spots are most likely due to differences in the input material. However, including differentially treated NB4 cells in the PRISMA screen facilitated the detection of several C/EBPα interactors that are regulated by ATRA or TPA treatment. The transcriptional coactivator endothelial differentiation related factor 1 (EDF1) regulates DNA-binding activity of the bZIP transcription factors ATF1, ATF2, CREB1 (Miotto and Struhl, 2006). EDF1 was down-regulated by TPA treatment and displayed binding to PRISMA peptides corresponding to the CR6 region. The barrier to autointegration factor 1 (BANF1) plays a role in chromatin organisation and was up-regulated by ATRA and TPA treatment. BANF1 bound to peptides corresponding to C/EBPα CR3,4 and bZIP domain.

Figure 15: Differential interactors of TPA and ATRA treated NB4 cells in PRISMA. SILAC ratios of significant interactors across PRISMA peptides (ordered from C/EBPα N to C terminus) are displayed. The first column indicates the SILAC ratio of the respective protein in the input material **A**: Most differential interactors in ATRA treated cells. **B**: Most differential interactors in TPA treated cells.


A
ATRA(H)/control(L) SILAC ratios of interactors across PRISMA peptides
log2(H/L)
4
3
2
1
0
−1
−2
SFPQ
HIST1H1B
PHGDH
TOP1
DDX3X
CAPRIN1
GTF2F2
GTF2F1
PARP1
TOP2A
TOP2B
UBTF
NCF1;NCF1B
PRAM1
BANF1
B
TPA(M)/control(L) SILAC ratios of interactors across PRISMA peptides
log2(M/L)
4
3
2
1
0
−1
−2
IGF2BP2
SFPQ
CAPRIN1
TTC9C
DCXR
TPX2
STK10
HMGN1
BANF1
BIN2
PARP1
UBTF
PIN4
EDF1
DNMT1
QLPESSLVK.me2.AMGNCA
QLPESSLVKAMGNCA
QLSRELDTLRGIFRQ
RLRKRVEQLSRELDT
ELTSDNDHLRKRVEQ
ETQQKVLELTSDNDR
KAKQRNVETQQKVLE
AVRKSRDKAKQRNVE
RRERNNIAVRKSRDK
NSNEYRVRRERNNIA
AKKSVDKNSNEYRVR
GGSGAGKAKKSVDK.ac.N
GGSGAGKAKKSVDKN
HPDLR.me2_sym.ASGGSGAGKA
HPDLRAS.phos.GGSGAGKA
HPDLR.me2_asym.ASGGSGAGKA
HPDLRASGGSGAGKA
LKGLGAAHPDLR.me2_asym.ASG
LKGLGAAHPDLR.me2_sym.ASG
LKGLGAAHPDLRAS.phos.G
LK.ac.GLGAAHPDLRASG
LK.me2.GLGAAHPDLRASG
LKGLGAAHPDLRASG
LPGPGSALK.me2.GLGAAH
LPGPGSALK.ac.GLGAAH
LPGPGSALKGLGAAH
PALGAAGLPGPGSAL
VPS.phos.PHPAPALGAAGL
VPSPHPAPALGAAGL
PTPPPTPVPS.phos.PHPAP
PTPPPT.phos.PVPSPHPAP
PT.phos.PPPTPVPSPHPAP
PTPPPTPVPSPHPAP
MHLQPGHPTPPPT.phos.PV
MHLQPGHPT.phos.PPPTPV
MHLQPGHPTPPPTPV
AHCGQTTMHLQPGHP
PHLQFQIAHCGQTTM
PPAHLAAPHLQFQIA
S.phos.HPHPHPPPAHLAAP
SHPHPHPPPAHLAAP
PPPPPPPS.phos.HPHPHPP
PPPPPPPSHPHPHPP
AGLFPYQPPPPPPPS
EAK.me2.QLALAGLFPYQP
EAK.ac.QLALAGLFPYQP
EAKQLALAGLFPYQP
QEPR.citr.EEDEAKQLALA
QEPR.me2_asym.EEDEAKQLALA
QEPREEDEAK.ac.QLALA
QEPREEDEAK.me2.QLALA
LRPLVIKQEPR.citr.EEDE
LRPLVIKQEPR.me2_asym.EEDE
LRPLVIK.ac.QEPREEDE
LRPLVIK.me2.QEPREEDE
LRPLVIKQEPR.me2_sym.EEDE
LR.me2_asym.PLVIKQEPREEDE
LR.me2_sym.PLVIKQEPREEDE
LRPLVIKQEPREEDE
ERVGAPALRPLVIK.ac.Q
ERVGAPALRPLVIK.me2.Q
ERVGAPALR.me2_asym.PLVIKQ
ERVGAPALR.me2_sym.PLVIKQ
ER.me2_asym.VGAPALRPLVIKQ
ERVGAPALRPLVIKQ
GRLEPLYER.me2_sym.VGAPAL
GRLEPLYER.me2_asym.VGAPAL
GR.me2_sym.LEPLYERVGAPAL
GR.me2_asym.LEPLYERVGAPAL
GRLEPLYERVGAPAL
AAAGYLDGR.me2_sym.LEPLYE
AAAGYLDGR.me2_asym.LEPLYE
AAAGYLDGRLEPLYE
PPPGYGCAAAGYLDG
M.Nterm.ac.PGGAHGPPPGYGCA
MPGGAHGPPPGYGCA
AGPGGAVMPGGAHGP
FDYPGAPAGPGGAVM
GGGGGGGDFDYPGAPA
K.ac.AAVGPTGGGGGGDF
SRQQEK.me2.AKAAVGPTG
SRQQEKAK.ac.AAVGPTG
SRQQEK.ac.AKAAVGPTG
SR.me2_sym.QQEKAKAAVGPTG
SR.me2_asym.QQEKAKAAVGPTG
SRQQEKAK.me2.AAVGPTG
SRQQEKAKAAVGPTG
LADLFQHSRQQEK.me2.AK
LADLFQHSR.me2_sym.QQEKAK
LADLFQHSR.me2_asym.QQEKAK
LADLFQHSRQQEK.ac.AK
LADLFQHSRQQEKAK
AAFNDEFLADLFQHS
ISAYIDPAAFNDEFL
EHETSIDISAYIDPA
EPLGGICEHETSIDI
PAPPAAPEPLGGICE
GAGPAQPPAPPAAPE
AAFGFPR.me2_asym.GAGPAQPP
AAFGFPR.me2_sym.GAGPAQPP
AAFGFPRGAGPAQPP
PPHAPSSAAFGFPR.me2_asym.G
PPHAPSSAAFGFPR.me2_sym.G
PPHAPSSAAFGFPRG
MSSHLQS.phos.PPHAPSSA
MSSHLQSPPHAPSSA
EAEPR.me2_asym.PPMSSHLQSP
EAEPRPPMSSHLQS.phos.P
EAEPR.me2_sym.PPMSSHLQSP
EAEPRPPMSSHLQSP
MESADFYEAEPR.me2_asym.PPM
MESADFYEAEPR.me2_sym.PPM
M.Nterm.ac.ESADFYEAEPRPPM
MESADFYEAEPRPPM
ratio in input material
bZIP
CR7
CR6
CR1L
CR5
CR4
CR3
CR2

3.3.1.3. PTMs modulate peptide-protein interactions in PRISMA

C/EBP proteins are heavily post-translationally modified, contributing to the functional plasticity of these proteins and possibly influencing protein interactions. The PRISMA screen included 40 post-translational modifications that were incorporated into 70 singly modified peptides. Modifications in the bZIP domain were not considered since they most likely influence DNA binding rather than PPIs. Most methyltransferases are not only capable of monomethylation, but also dimethylation, therefore only dimethylation was included in the screen. The influence of PTMs on protein interaction in PRISMA was evaluated by a fold change cut-off described in detail in the method section. The barplot in **Figure 16A** depicts the number of interactions detected by PRISMA for the unmodified peptide (black bars) and the number of interactions that were enhanced (red bars) or decreased (blue bars) by a PTM of the respective peptide. In **Figure 16B** selected examples of PTM modulated interactions are displayed.

Methylation of the arginine 12 residue enhanced interaction with the Mediator subunit MED16 and decreased interaction with the cyclin-dependent kinase CDK13. Together with its paralog CDK12, CDK13 has been described to be involved in RNA processing and gene regulation (Liang et al., 2015). The Protein arginine N-methyltransferase 1 (PRMT1) was found to bind stronger to an unmodified peptide spanning R35 compared to the dimethylated version, suggesting that PRMT1 might methylate R35. The K161 residue has been previously reported to be sumoylated by Ubc9 which ultimately led to the degradation of C/EBPα (Geletu et al., 2007). Compared to peptides containing a modification at K161, the ubiquitin conjugating enzyme UBE2E1 (also sometimes referred to as UBCH6) bound stronger to the unmodified counterpart. This indicates that UBE2E1 might be involved in ubiquitinating C/EBPα at position 161 and that other K161 modifications may protect C/EBPα from ubiquitination and subsequent proteasomal degradation. EDF1 was down-regulated by TPA treatment (**Figure 15B**), and interacted with the K161 acetylated C/EBPα peptide but not the unmodified counterpart. PRISMA suggested increased binding of SMARCE1 and EP400 to the R142 methylated peptides as compared to the unmodified counterpart. SMARCE1 is a subunit of the nucleosome remodelling complex SWI/SNF while EP400 is part of the NuA4 histone acetyltransferase complex. Both complexes are involved in chromatin reorganisation during haematopoietic development and EP400 knockout in mice leads to defects in embryonic and adult bone marrow haematopoiesis (Prasad et al., 2015; Ueda et al., 2007). The transcriptional repressor THAP11 is down-regulated during erythroid differentiation and overexpression of he protein inhibited differentiation of the erythroid K562 cell line (Kong et al., 2014). In PRISMA, THAP11 was interacting only with a C/EBPα peptide

containing an arginine citrullination at R165. Since many of the PTM dependent C/EBPα interactors identified in PRISMA are connected to haematopoiesis and differentiation, further studies will be required to evaluate the biological function of these novel interactions *in vivo*.

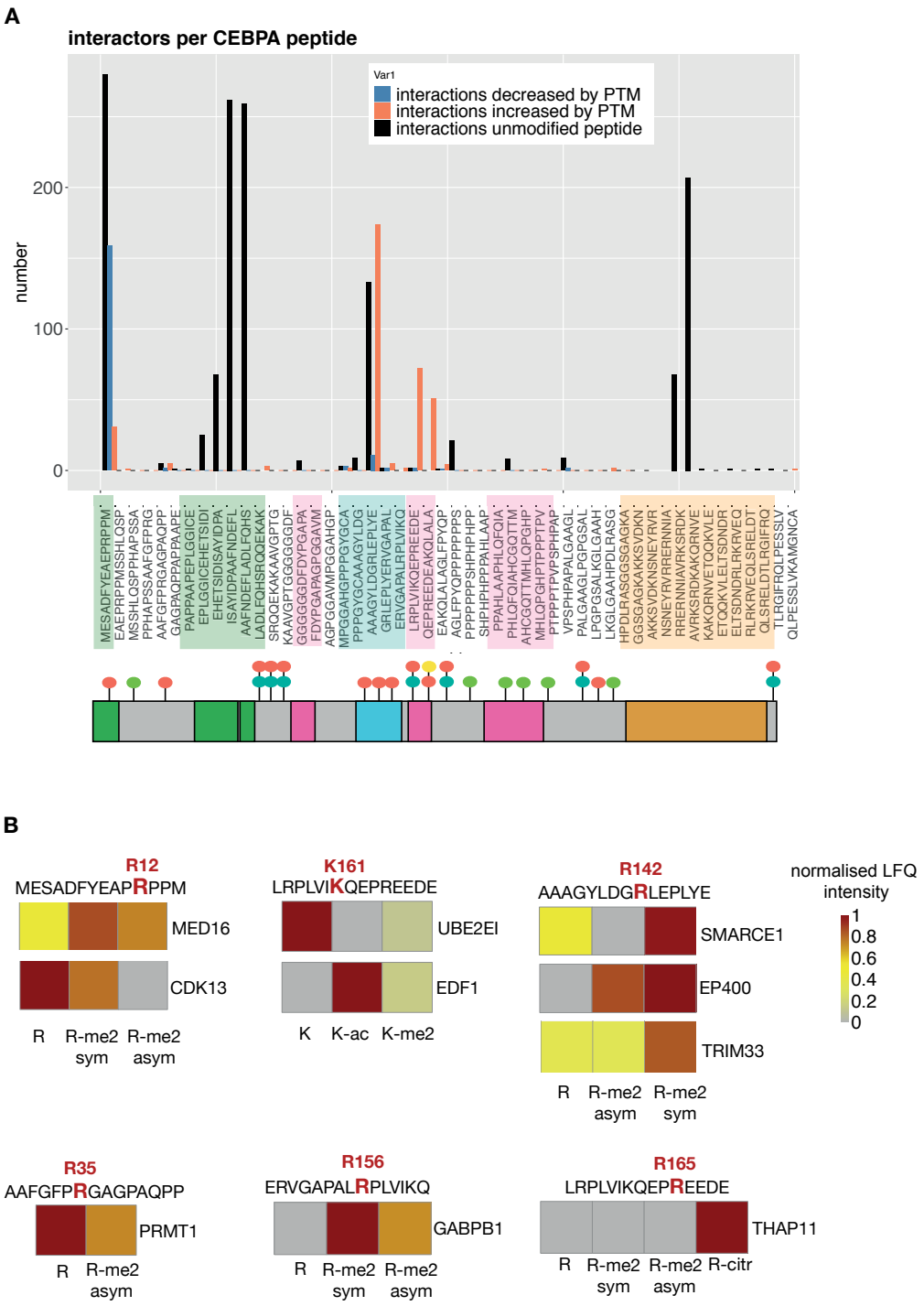


Figure 16: PRISMA detected PTM dependencies of C/EBPα protein interactions. **A:** Barplot depicts the number of interactions per PRISMA C/EBPα peptide ordered from N- to C-terminus. Black bars represent the number of interactions of the unmodified peptide. Red and blue bars represent the number of interactions that were increased or decreased by a modification (Cut-offs: positive enrichment against other peptides spots with an FDR <10% and log_2(modified/unmodified) in both replicates > 0.75 or < -0.75) **B:** Selected PTM modulated C/EBPα interactions as detected by PRISMA are displayed, color code corresponds to normalised LFQ intensities.

3.3.2. C/EBPα BioID

BioID experiments in NB4 cells were employed to validate PRISMA data and to gain a more detailed insight into the C/EBPα isoform-specific interactome. For this purpose, stable inducible NB4 cell lines were generated that expressed either C/EBPα-BioID or the BioID tag alone (control) (**Figure 17A**). As an additional control sample, C/EBPα-BioID NB4 cells not treated with doxycycline were included in the experiment. Induction of BioID fusion proteins and successful biotinylation was confirmed by western blotting **(Figure 17B**). Experiments were performed in quadruplicates employing the label free quantification (LFQ) algorithm from MaxQuant (Cox et al., 2014). Pearson correlation of replicates was 0.9 or higher, indicating high reproducibility of BioID experiments (**Figure 17C**). BioID identified 354 high confidence C/EBPα proximity interactors in NB4 cells (two sided t-test, FDR < 5%, log_2(enrichment) > 1 against both controls). Among the most enriched proteins were several transcription factors of the C/EBP and ATF families representing known heterodimerisation partners of C/EBPα (McKnight, 1991), confirming successful proximity labelling and enrichment of interactors (**Figure 17D**). A ranked table containing all C/EBPα interactors identified by BioID experiments is available in the supplements (**Supplemental Table 1**).

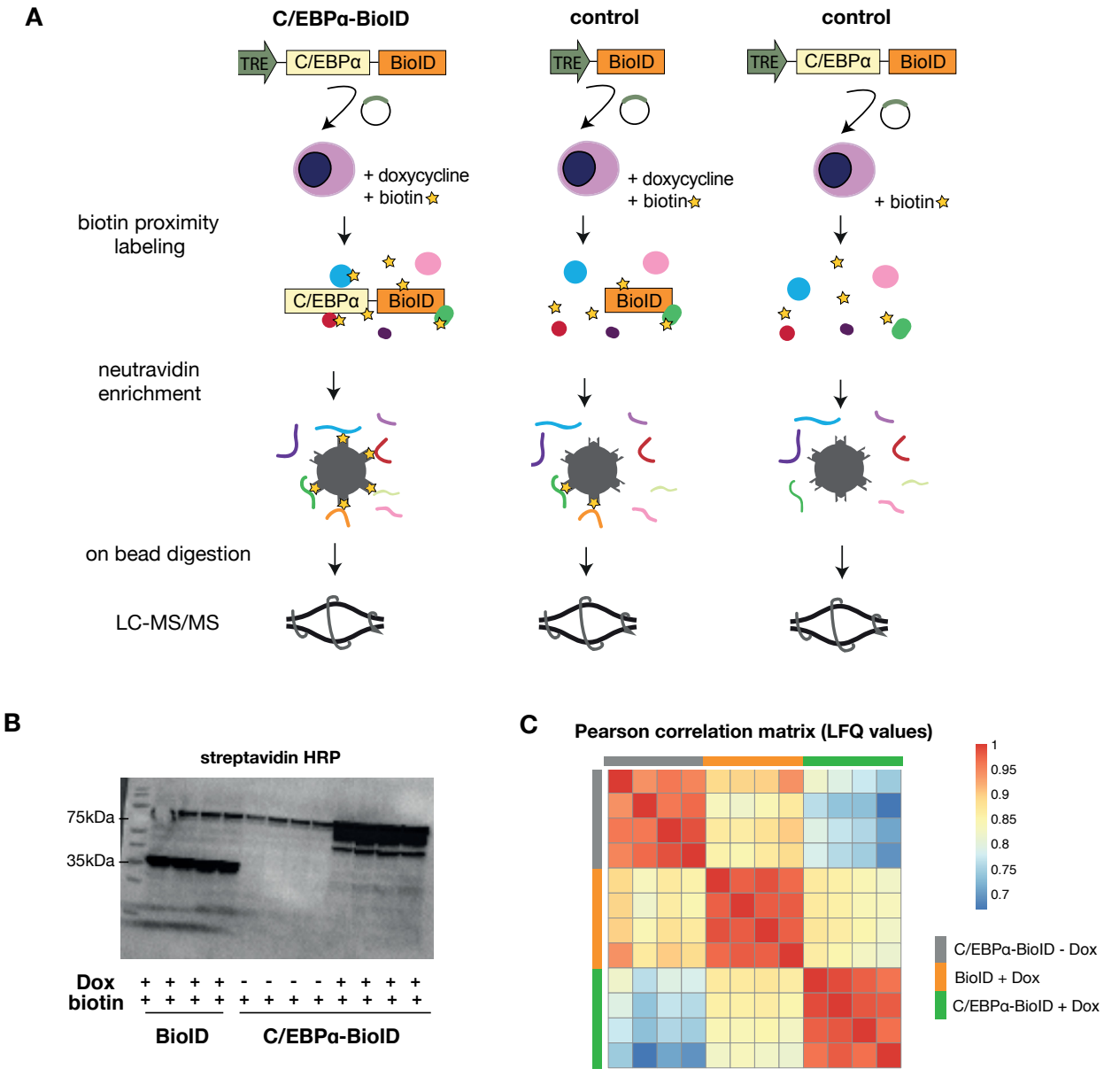
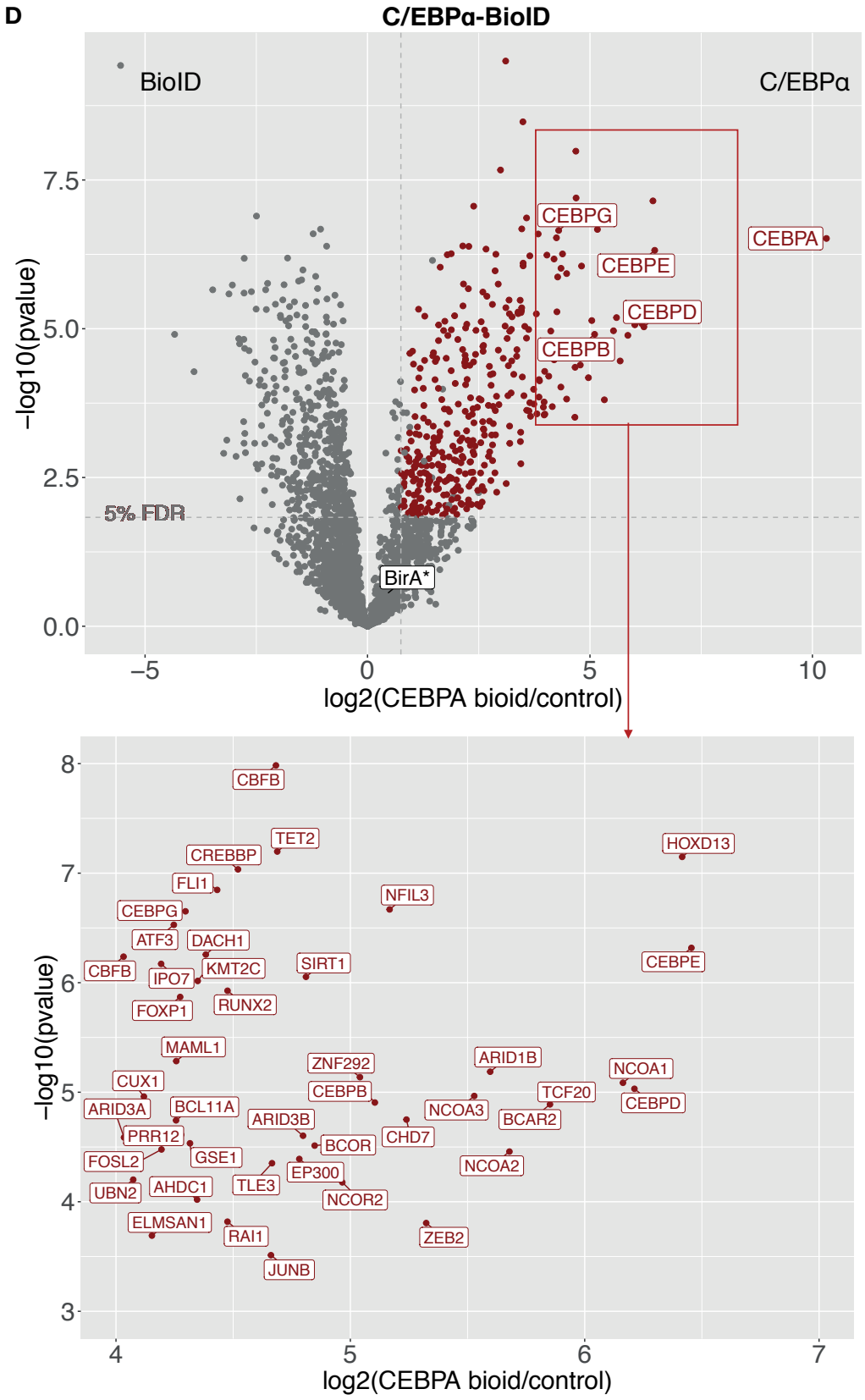


Figure 17: BioID detects C/EBPα interactors in live NB4 cells.
A: Stable NB4 cell lines inducibly expressing C/EBPα-BioID under the control of a tet-responsive element were created. Upon induction of the construct and addition of biotin, proteins in close proximity to the fusion protein are biotinylated and then subsequently enriched with neutravidin beads. Following stringent washing steps, bound proteins are digested on bead. Proteins were and identified and quantified with label free LC MS/MS. As controls cells expressing only the biotin ligase and cells not treated with doxycycline were employed. **B:** Western blotting and detection with streptavidin-HRP confirmed successful induction of BioID and biotinylation **C**: Pearson correlation matrix of BioID pull-downs (LFQ values). Each experiment contained 4 biological replicates. **D:** Enrichment of proteins in C/EBPα-BioID vs BioID control (x axis) is plotted against their $-\log_{10}$(p-values). C/EBP factors and biotin ligase are indicated. Plot at the bottom represents a zoom in as indicated of the upper plot.


D
C/EBPα-BioID
BioID
C/EBPα
−log10(pvalue)
7.5
5.0
2.5
0.0
5% FDR
BirA*
CEBPG
CEBPA
CEBPE
CEBPD
CEBPB
−5
0
5
10
log2(CEBPA bioid/control)
8
7
6
5
4
3
CBFB
TET2
CREBBP
FLI1
CEBPG
ATF3
DACH1
KMT2C
IPO7
FOXP1
RUNX2
SIRT1
NFIL3
HOXD13
CEBPE
MAML1
CUX1
ARID3A
BCL11A
PRR12
FOSL2
GSE1
UBN2
AHDC1
ELMSAN1
RAI1
JUNB
TLE3
EP300
NCOR2
ZNF292
CEBPB
ARID3B
BCOR
CHD7
ZEB2
ARID1B
NCOA3
NCOA2
BCAR2
TCF20
NCOA1
CEBPD
4
5
6
7
log2(CEBPA bioid/control)

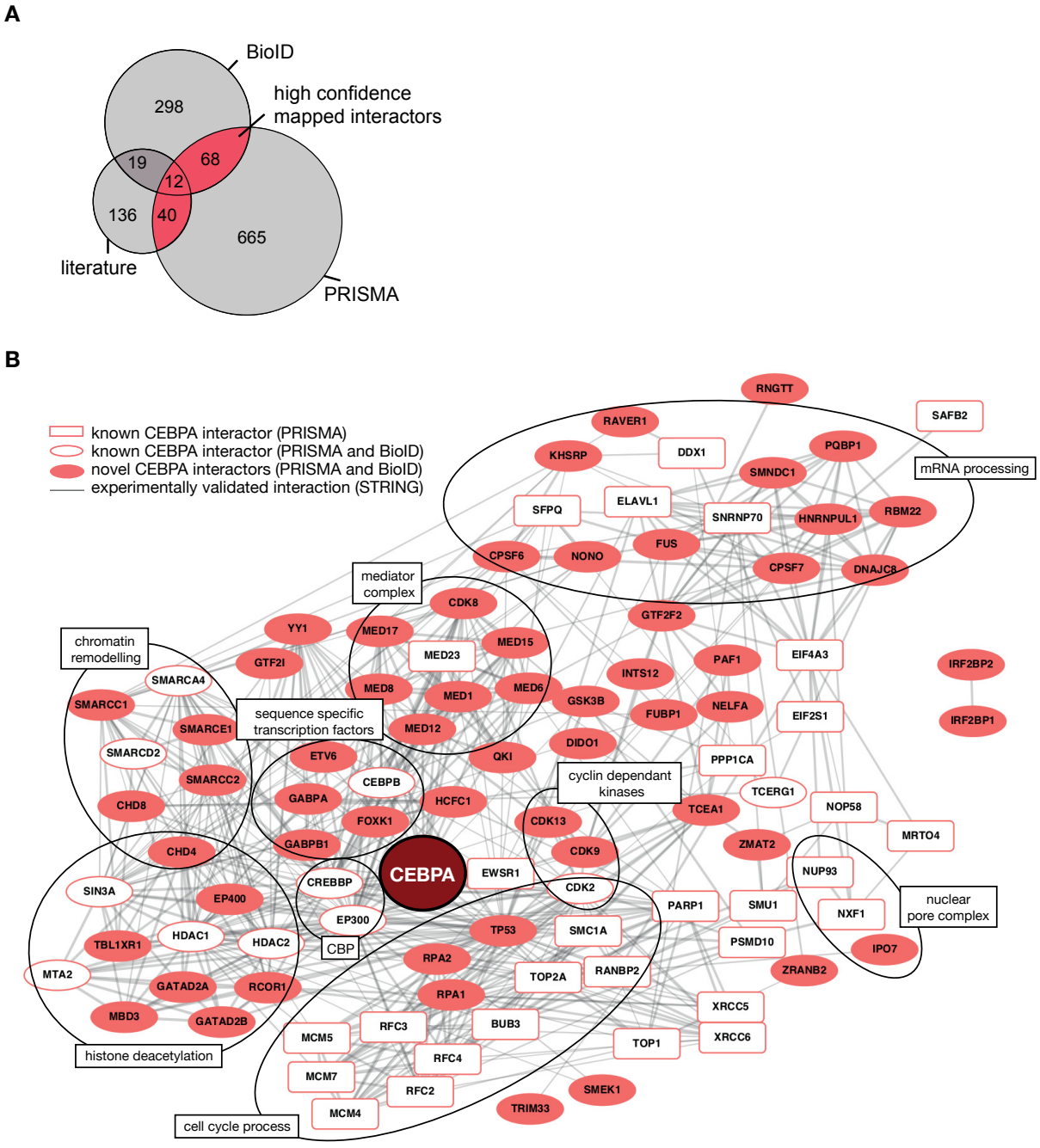
3.3.3. Overlap of C/EBPα PRISMA and BioID

The C/EBPα interactomes derived from BioID and PRISMA were compared with published C/EBPα interactors (BioGrid (Chatr-Aryamontri et al., 2017) and STRING (Szklarczyk et al., 2015) databases, Grebien et al., 2016). In total, 80 proteins overlap between the PRISMA- and BioID-derived C/EBPα interactomes, of which 12 are previously identified interactors **(Figure 18A)**. The 120 PRISMA interactors that were validated by either BioID or databases (40 proteins) make up a subset of high confidence C/EBPα interactors that can be depicted across the linear C/EBPα sequence and PTM sites (**Supplemental Figure 2**). These 120 interactors show high connectivity according to experimentally validated interactions listed in the STRING database (**Figure 18B**). They can be separated into several functional groups and protein complexes like the MED complex, sequence specific transcription factors, mRNA processing proteins, histone deacetylases as well as other chromatin remodelling enzymes. Known interactors are indicated as white nodes in the network while coloured nodes represent novel C/EBPα interactors identified for the first time in this study. Novel interactors include the transcription factors GABPA and GABPB1 that are part of the tetrameric transcription factor complex GABP. GABP is required for myeloid differentiation (Yang et al., 2011) and could be a specific interactor of C/EBPα in myeloid cells. Additionally to the 120 C/EBPα interactors that were significant in both PRISMA and BioID, another 93 BioID C/EBPα interactors were detected but did not pass the significance threshold of 10% FDR in PRISMA. Their binding profile across C/EBPα PRISMA peptides is depicted in Supplemental Figure 3.

GO term enrichment of the validated interactors of each conserved C/EBPα region revealed that individual CRs are connected to distinct functional roles of C/EBPα, as shown **in Figure 18C**. CR 3/4 contains most of the significantly enriched GO terms, suggesting the importance of this core transactivating region to all P42-C/EBPα functions. No GO terms were found enriched with CR7 derived peptides, although several proteins like the transcriptional repressor YY1 and HTATSF1 interacted with this region, suggesting its functional heterogeneity. Most interactors and enriched GO terms map to CR2 and CR3/4, which are unique to the N-terminal part of P42-C/EBPα, and CR1L, which constitutes the N-terminus of the truncated isoform P30. The PRISMA data predicts that P30 can still function to recruit major components of the transcriptional and epigenetic machinery albeit with lower efficiency compared to full length C/EBPα.

As expected, a number of C/EBPα interactions are detected in only one of the two datasets (**Figure 18A)**. Comparison of intensity based absolute quantification values (iBAQ) (Schwanhäusser et al., 2011) in the NB4 nuclear proteome of PRISMA

and BioID interactors revealed a preference of PRISMA for more abundant interactors (**Figure 19**). The majority of proteins that were detected in only one of the two datasets are direct interactors of the 120 C/EBPα interactors depicted in Figure 18B. Over 70% of the C/EBPα interactors detected only in PRISMA (468 proteins, **Supplemental Figure 4**) or BioID (278 proteins, **Supplemental Figure 5**), are connected to the validated interactors by at least one experimentally validated interaction deposited in the STRING PPI database (edges in Supplemental Figure 4 and 5).



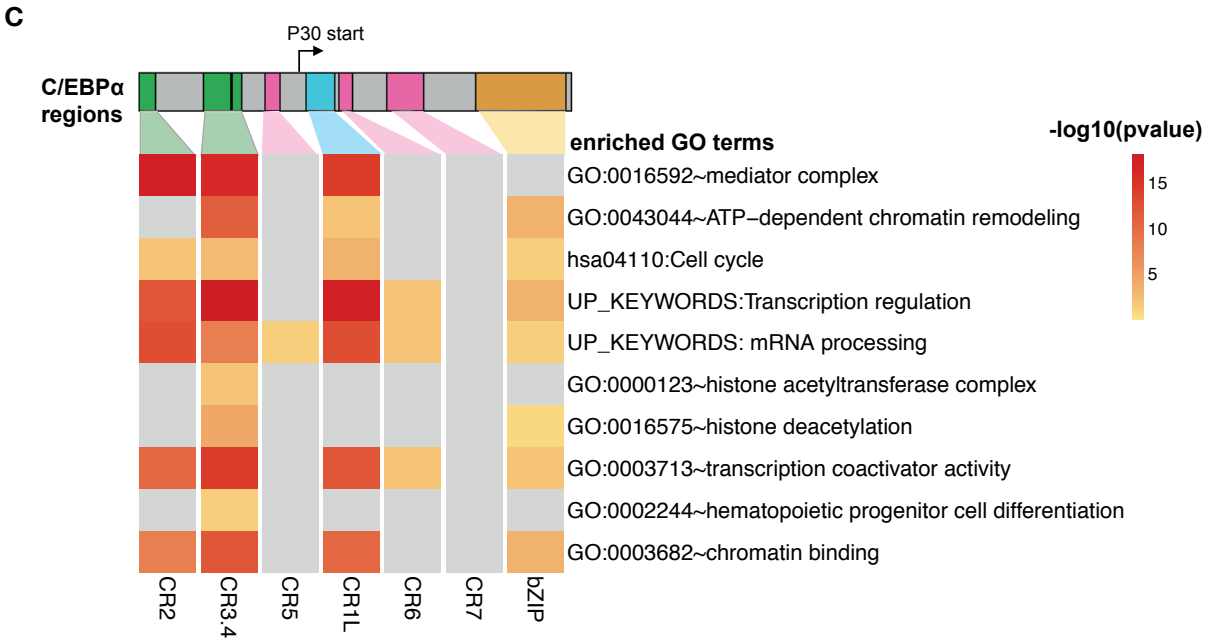


Figure 18: Integration of BioID and PRISMA data validates linear C/EBPα interactors.
A: Venn diagram depicting the overlap of BioID, PRISMA and published C/EBPα interactors deposited in the BioGRID and STRING databases and Grebien et al., 2016. **B:** Interaction network of validated C/EBPα PRISMA interactors (overlap between PRISMA and BioID or PRISMA and literature) visualised with Cytoscape. Edges represent experimentally validated interactions retrieved from the STRING database. Novel C/EBPα interactors are depicted as coloured nodes, known interactors are depicted as white nodes. Interactors identified only by PRISMA are displayed with a rectangular outline, round outline indicate interactors identified with PRISMA and BioID. Interactors are grouped by functional annotation. Interactors not connected by any edges were removed from the plot (12 proteins). **C**: Interactors mapped with PRISMA to conserved C/EBPα regions were subjected to GO term analysis with DAVID tool. Informative significant GO terms (p-value < 0.05) are displayed. Grey indicates no significant enrichment.

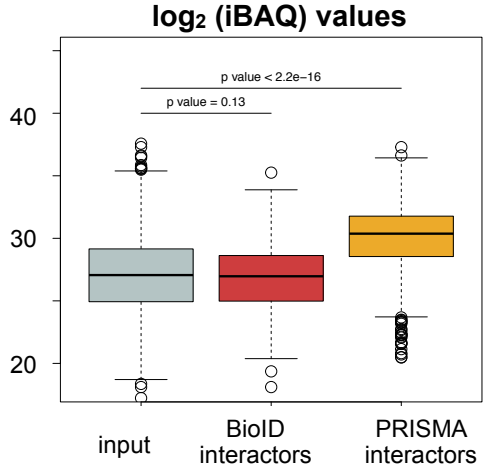


Figure 19: IBAQ values of C/EBPα interactors detected by PRISMA or BioID.
Boxplots depict log_2 of intensity based absolute quantification (iBAQ) values of the detected interactors in the PRISMA input material (NB4 nuclear extract). Indicated p-values were calculated with a Welch test.

3.3.3.1. C/EBPα methylation in CR1L enhances interaction with the SWI/SNF complex

PRISMA suggested increased binding of the SMARCE1 subunit of the nucleosome remodelling complex SWI/SNF to the R142 methylated peptides as compared to the unmodified counterpart. Other subunits of SWI/SNF followed the same trend but differential binding to the methylated peptide spanning R142 scored below the statistical significance threshold (**Figure 20A)**. The methylation-enhanced interaction of C/EBPα with SMARCE1 was confirmed with BioID experiments with a methylation mimicking mutant (residues R142/149/156 converted to triple L, depicted as L-mutant in Figure 20). SMARCE1 and three additional SWI/SNF subunits (ARID1A, ARID1B, ARID2) were found significantly enriched in the L-mutant, as compared to WT C/EBPα-BioID **(Figure 20B)**. BioID with L-mutant-C/EBPα also verified the methylation dependent interaction with the E3 ubiquitin ligase TRIM33 that was detected by PRISMA. The NURD complex component GATAD2A was significantly enriched in L-mutant C/EBPα-BioID in comparison to the wild type protein. In PRISMA, GATAD2A bound with higher intensity to the R142 methylated peptide, although it failed the initial significance threshold **(Figure 20C)**. Several Myb-Muvb/DREAM complex members (LIN9, LIN37, MYBL2) were identified as L-mutant-C/EBPα specific interactors by BioID but were not detected in PRISMA.

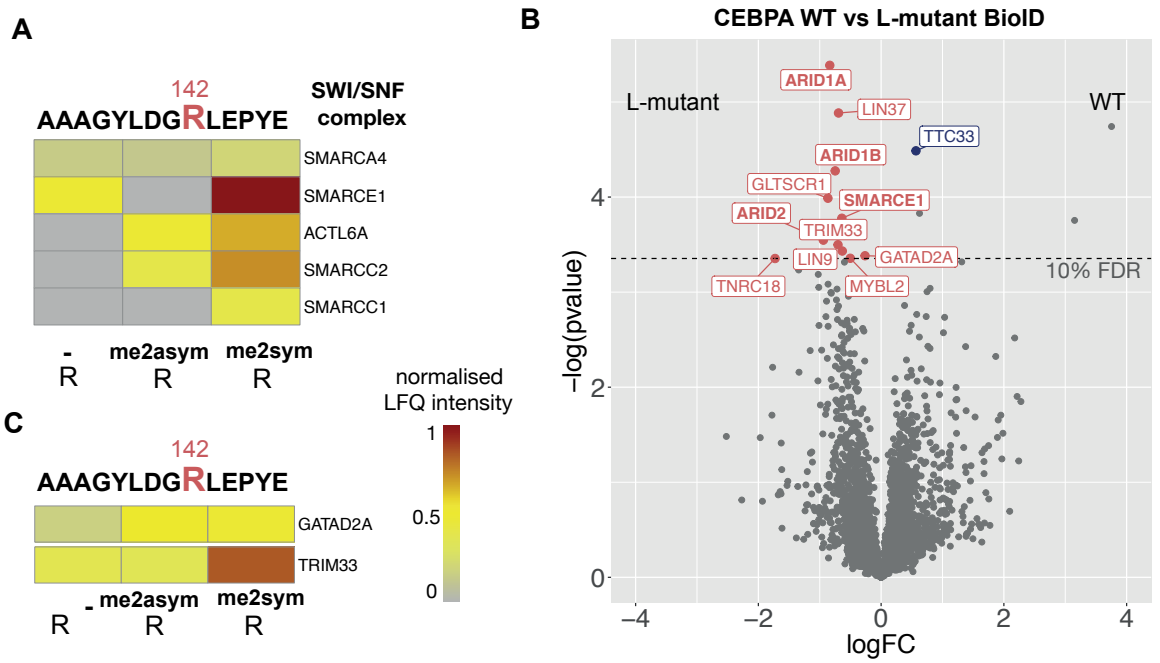


Figure 20: SMARCE1 interaction with C/EBPα CR1L is methylation dependent
A: Heatmap displaying normalised LFQ intensities of SMARCE1 and other SWI/SNF complex member to C/EBPα PRISMA peptides spanning R142 in CR1L. **B**: BioID pull-downs in NB4 cells with wild type P30-C/EBPα (WT) and a methylation-mimicking mutant (R142/R147/R154->L; L-mutant). Proteins passing the significance cut-offs against the BioID control and differentially binding to WT or L-mutant C/EBPα are indicated. SWI/SNF complex members are marked in bold letters. **C:** PRISMA binding profiles of additional L-mutant C/EBPα specific interactors.

3.3.4. Comparison of interactome data from C/EBPα and C/EBPβ PRISMA screens

The C/EBP transcription factor family consists of six members that all contain a bZIP domain in the C-terminus. In addition, C/EBPα,β,δ,ε show local similarities between their N-terminal conserved regions. The sequence alignment and homologue conserved regions of C/EBPα,β is depicted in **Figure 21A,B**. Comparison of the PRISMA data from C/EBPα and previously published PRISMA data from C/EBPβ (Dittmar et al., 2019) revealed that homologous regions also share a number of interactions **(Figure 21C)**. The bZIP domain, which is the most conserved between all C/EBP transcription factors, also displays the largest interactor overlap between C/EBPα and C/EBPβ while the regions CR1L (C/EBPα) and CR1 (C/EBPβ) that are structurally distinct but share functional similarities, also share some common interactors. Despite differences in the experimental setup between both C/EBPα and C/EBPβ PRISMA experiments, the mediator complex was found to bind to the same homologue CRs in both datasets **(Figure 21D)**.

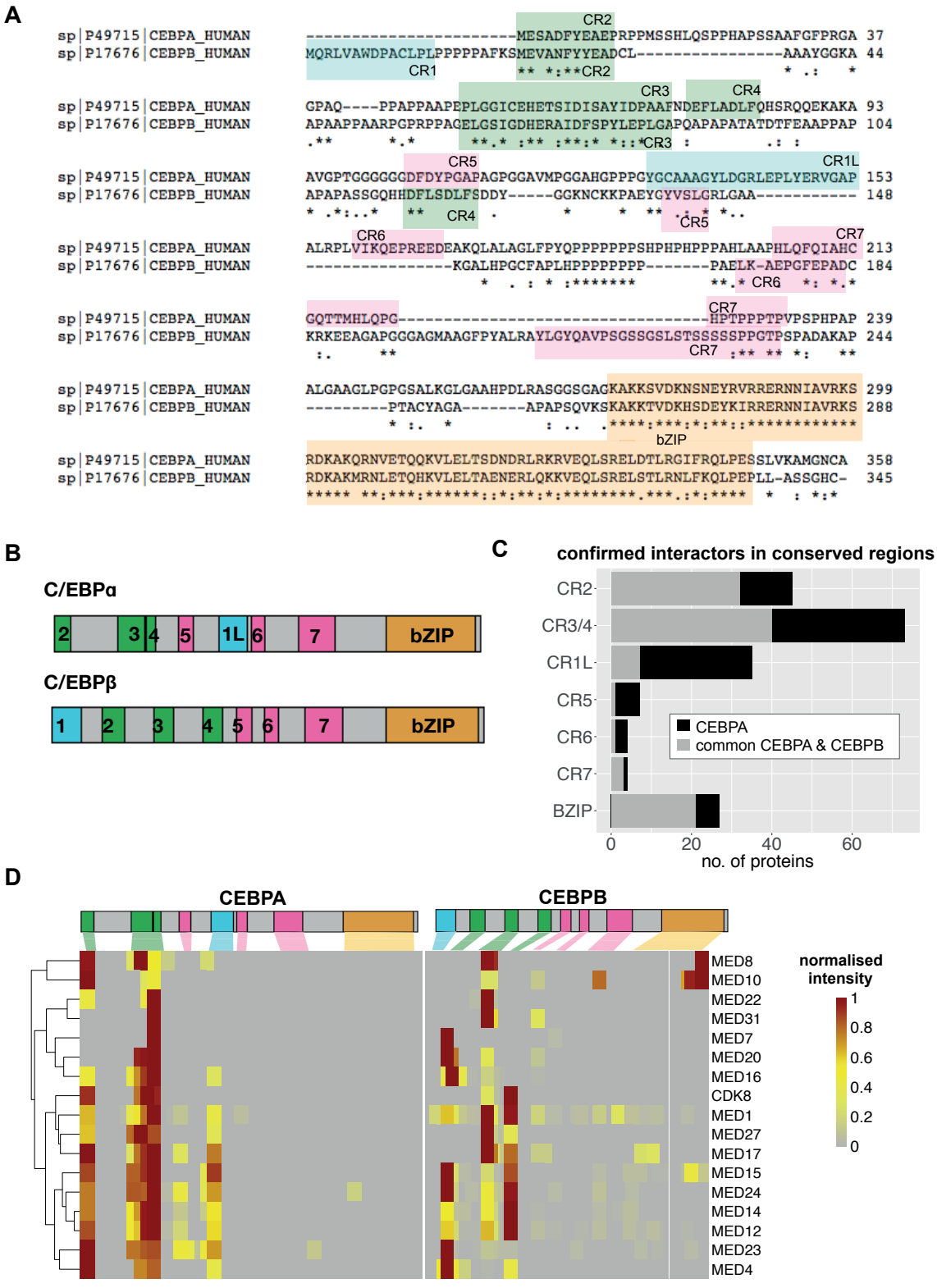


Figure 21: C/EBPα and C/EBPβ share interactors in homologous regions.
A: Sequence alignment of human C/EBPα and C/EBPβ **B**: Conserved regions in C/EBPα and C/EBPβ **C:** Number of validated C/EBPα interactors per conserved region in C/EBPα (black bars). Grey bars represent interactors that were also binding to homolog regions in C/EBPβ. **D:** Extracted binding profile of Mediator complex subunits binding to C/EBPα(left) and C/EBPβ (right). Annotation bar on top indicates conserved regions.

3.3.5. The isoform-specific C/EBPα interactome

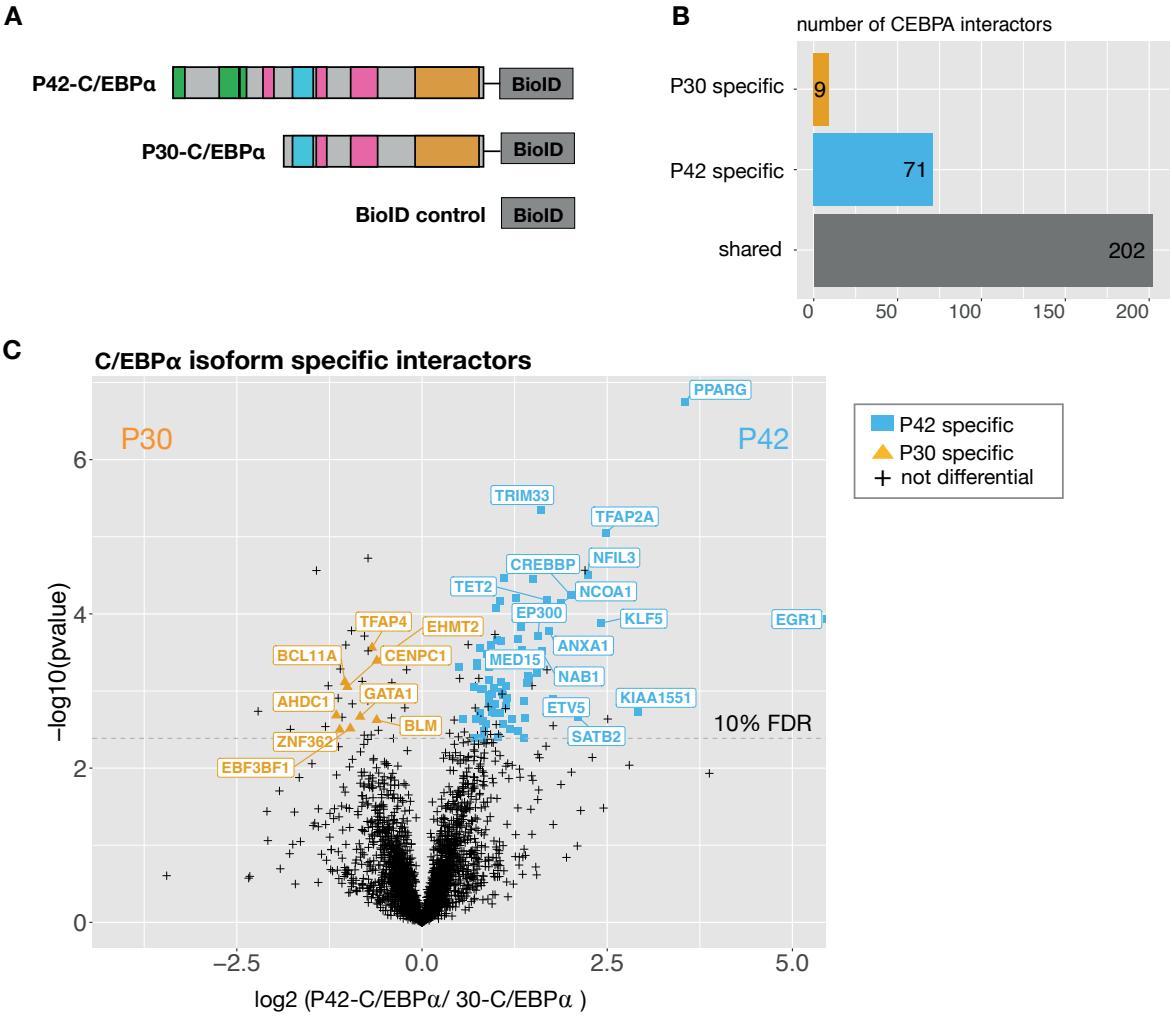
To further explore the C/EBPα isoform-specific interactome, P42- and P30-C/EBPα were expressed as BioID fusion proteins in NB4 cells **(Figure 22A)**. The vast majority of interactors were found to interact with both C/EBPα isoforms (5% FDR, 2-fold enrichment against both controls), while the direct comparison of P42 and P30 revealed 80 isoform-specific interactors (10% FDR) **(Figure 22B,C, Supplemental Table 2)**. This is in line with the results obtained by PRISMA, suggesting multi-valency of distinct peptides in several C/EBPα CRs, including CR1L as part of P30-C/EBPα. Comparing the quantitative enrichment of interactors in P42- and P30-C/EBPα-BioID against the BioID control further suggested that interactors were pulled down with similar efficiency with both C/EBPα isoforms **(Figure 22D)**.

Among others, the erythroid master regulator GATA1, the SWI/SNF associated transcription factor BCL11A and the proliferation regulating transcription factor TFAP4 were identified as P30-C/EBPα specific interactors in NB4 cells. Most P30-C/EBPα specific interactors also interacted with P42-C/EBPα but displayed lower affinity for the P42 isoform. In contrast, a subset of P42-C/EBPα interactors exclusively interacted with the P42-C/EBPα isoform and were not part of the P30-C/EBPα interactome. These P42 exclusive interactors include the transcription factor early growth response protein 1 (EGR1) that is involved myeloid differentiation (Krishnaraju et al., 2001; Nguyen et al., 1993) and the peroxisome proliferator-activated receptor gamma (PPARG) that is a master regulator of adipogenesis and mediator of macrophage development (Tontonoz et al., 1994, Chinetti et al., 1998; Lefterova et al., 2014).

PRISMA data, as shown in **Figure 22E,** confirmed the isoform-specificity of protein interactions observed in BioID experiments. P42-C/EBPα specific interactors identified with proximity labelling also showed higher affinity for PRISMA peptides that correspond to the unique part of P42-C/EBPα, as compared to peptides that are shared between both C/EBPα isoforms. In contrast, interactors that were specific for P30 or shared between both C/EBPα isoforms in BioID experiments displayed stronger binding to PRISMA peptides that were derived from P42/P30 shared regions.

The results from BioID and PRISMA are in contrast with findings from a previous study suggesting that the interactome of the two C/EBPα isoforms were largely different with the MLL subunit WDR5 as a differential interactor of P30-C/EBPα (Grebien et al., 2016). Moreover, while the BioID experiments presented here confirmed binding of WDR5 and other MLL subunits to C/EBPα and previously also to C/EBPβ (Dittmar et al., 2019), differential binding of WDR5 to P30- compared to P42-C/EBPα was not observed.

Proteins differentially interacting with the P30 isoform may pose a selective vulnerability of P30-expressing cells and a therapy target for C/EBPα mutated AML cases. CRISPR/Cas9 knockout study derived dependency scores of the P30-C/EBPα specific interactors in 18 different AML cell lines were extracted from the Depmap portal (Meyers et al., 2017; Tsherniak et al., 2017) **(Figure 22F).** As a reference, the dependency scores of the tumour suppressor TP53 and MYB oncogene are plotted on top of **Figure 22F**. Half of the AML cell lines tested (9 out of 18) were sensitive to TFAP4 knockout (threshold < –0.5) while two and one cell line tested are sensitive to GATA1 and BCL11A or BLM knockout respectively. The TFAP4 dependency of AML cell lines hints at therapeutic intervention possibilities of N-terminally mutated C/EBPα AML. The data from PRISMA and BioID experiments suggests that the interactomes of C/EBPα isoforms largely overlap and highlights P42/P30 specific interactions with lineage defining transcription factors that may fine-tune transcriptional outcome in haematopoietic cells



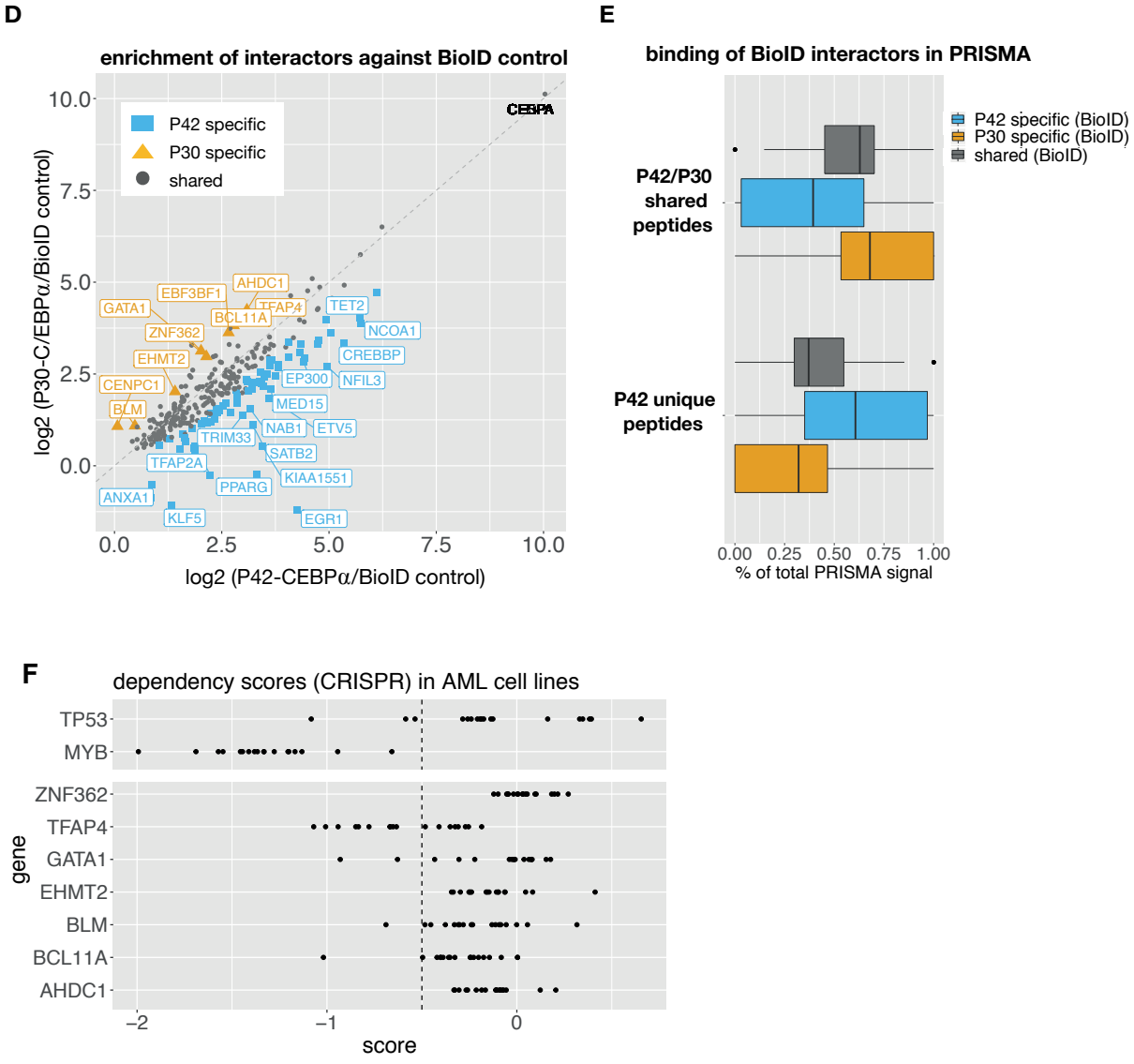


Figure 22: BioID detects C/EBPα isoform-specific protein interactions.
A: Stable pools of NB4 cells were engineered to express full length (P42) or the truncated (P30) C/EBPα isoform fused to BioID or the BioID tag alone. **B**: Number of C/EBPα isoform interactors identified by BioID experiments. **C**: Volcano plot directly comparing P42 and P30 BioID pull-downs (n = 4) with each other, ratio is plotted against $-\log_{10}$ (p-value). The significance threshold <10% FDR is indicated with a dotted line. **D**: Enrichment of C/EBPα interactors against BioID control; the ratio of P42-C/EBPα/BioID is plotted against the ratio of P30-C/EBPα/BioID. **E**: Distribution of isoform-specific interactors derived from BioID experiments in PRISMA data. **F**: Depmap dependency scores from CRISPR knockout experiments of P30 specific interactors in AML cell lines. Known tumour suppressor P53 and oncogene MYB are plotted on top as a reference.

3.4. Gene expression induced by expression of P30- and P42-C/EBPα in NB4 cells

In BioID experiments a number of C/EBPα isoform-specific interactors detected were other sequence specific transcription factors, hinting at either a direct physical interaction or very close proximity on chromatin. RNA expression profiling by microarray of NB4 cell lines (**Figure 23A**) revealed that expression of the two C/EBPα isoforms induced differential gene expression responses (**Figure 23B**). GSEA was performed to evaluate the regulation of other transcription factors targets and immune cell signatures by C/EBPα isoform expression in NB4 cells (**Figure 23C**). GSEA showed that a GATA1 signature was significantly enriched with P30 but not P42 expressing cells. This is of particular interest as one of the nine P30-C/EBPα specific interactors was the erythroid transcription factor GATA1. PPARG was identified as P42-C/EBPα specific interactor in NB4 cells. GSEA of microarray data demonstrated that published gene expression patterns of PPARG knockout macrophages (Rőszer et al., 2011) correlated with P42 but not P30 expressing cells. The transcription factor EGR1 specifically interacted with P42-C/EBPα, however only the gene expression signature of P30 but not P42-C/EBPα was enriched for EGR1 target genes. ChIP-sequencing during mouse liver regeneration has previously identified overlapping genomic binding sites for Egr1 and Cebpα,β (Jakobsen et al., 2013). The authors found that overlapping genomic regions bound by both Egr1 and Cebpα,β lacked an Egr1 target sequence and further experiments suggested that Egr1 can interact indirectly with DNA at Cebp cognate sequences through interaction with Cebps. In NB4 cells, EGR1 RNA expression levels were up-regulated by both P42- and P30-C/EBPα expression. In P42 expressing cells, EGR1 might be subsequently recruited to C/EBP target sequences by specific interaction with the C/EBPα N–terminus while this interaction is absent with P30-C/EBPα. Taken together, the data from interactome and gene expression profiling suggest that the specific interactions of C/EBPα isoforms with lineage defining transcription factors are implicated in co-regulation of target genes in the haematopoietic system.

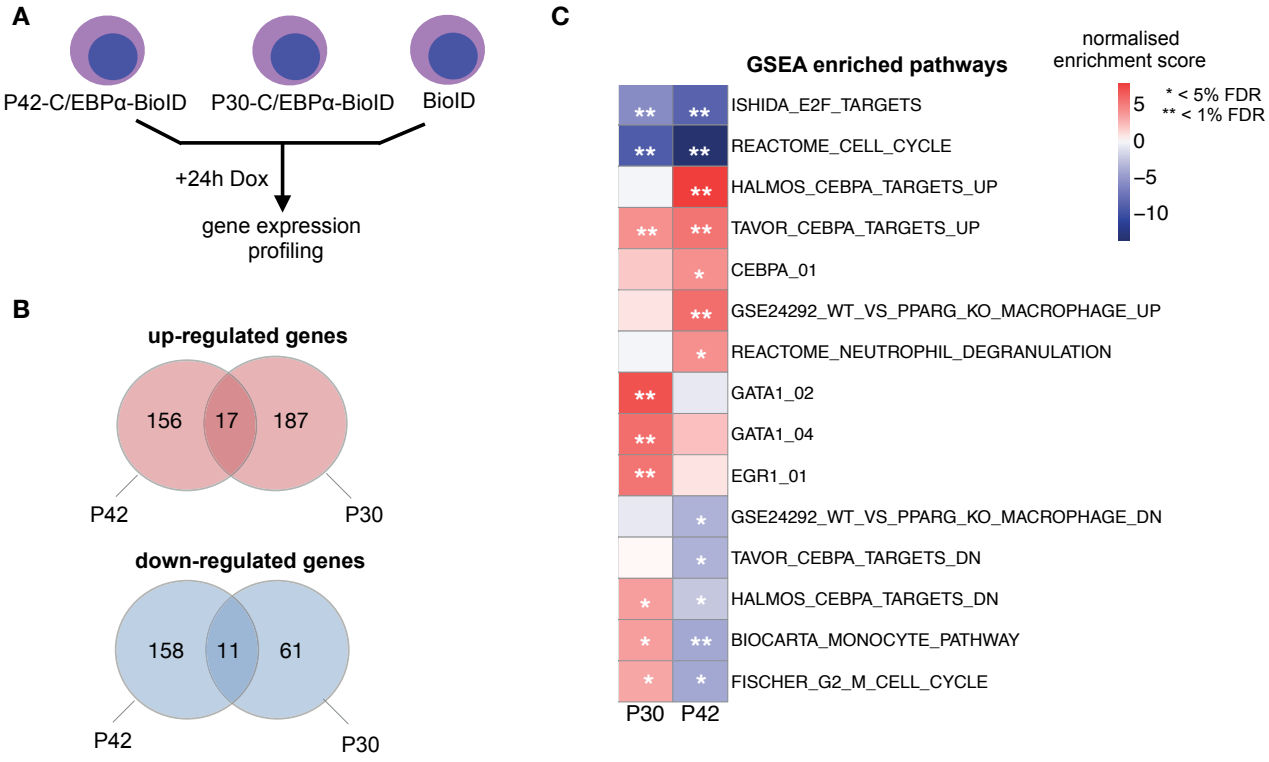


Figure 23: C/EBPα isoform expression induced differential gene expression In NB4 cells.
A: Stable pools of NB4 cells were engineered to express full length (P42) or the truncated C/EBPα isoform (P30) fused to BioID or BioID alone under the control of a doxycycline inducible promoter. Gene expression was induced for 24h and RNA expression analysed by microarray (n = 3). **B**: Overlap of up- and down-regulated genes between different C/EBPα isoforms (comparison to BioID expressing cells, < 5% FDR, abs(fold change) > 2) **C**: Gene set enrichment analysis (GSEA) of induced gene expression changes (relative to BioID control). Heatmap displays informative gene sets filtered for FDR <5%. Color scale corresponds to the normalised enrichment score (NES). * indicates an FDR <5% , ** indicates an FDR < 1%

4. Discussion

4.1. NB4 cells as a model system for myeloid differentiation and C/EBPα PPI studies

C/EBPα is a myeloid transcription factor and many of its known interactors, for example other C/EBP transcription factors, are only expressed in a specific subset of cells. Choosing the right model system for PPI studies is therefore crucial for the detection of such cell-type-specific interactions directly related to the biological function of C/EBPα. In the present study, NB4 cells were chosen as a model system for several reasons. First, NB4 are myeloid precursor cells and as such represent a cellular environment where many of C/EBPα cell-type-specific interactors are expressed. Like other promyelocytic leukaemia cells with a retinoic acid receptor fusion to PML (PML-RARA), NB4 cells can be induced to differentiate into granulocytes with ATRA. Differentiation of NB4 cells into monocytes/macrophages can be achieved by treatment with TPA or other chemical agents (Lanotte et al., 1991). This bi-lineage potential of NB4 cells may facilitate the detection of C/EBPα PPIs occurring in the context of myeloid differentiation. In contrast to primary cells, expansion of NB4 and obtaining enough material for a PRISMA screen - around 5mg of nuclear protein extract - was feasible. In addition, transduction of NB4 cells with retroviral constructs has been successful in the past (Darling et al., 2000) and enabled the generation of stable cell lines expressing C/EBPα-BioID fusion proteins.

Differentiation of NB4 cells into granulocytes and monocytes/macrophages, with ATRA or TPA respectively, was established. Kinetic changes of the transcriptome and the proteome were monitored over a time course spanning five (TPA) or seven (ATRA) time points. During NB4 differentiation vast changes on transcript and proteome level occur, and almost a third of all detected proteins and transcripts significantly changed in at least one of the analysed time points. Whether individual proteins and genes that are regulated in NB4 differentiation are drivers, or bystanders of myeloid differentiation remains to be evaluated. This data serves as a resource to facilitate the biological interpretation of future experiments and provides a basis to decipher regulators of myeloid differentiation. Among the most regulated proteins and transcripts are known factors of myeloid differentiation like the granulocytic transcription factor C/EBPε and JUN transcription factors. C/EBPα RNA and protein levels were slightly up-regulated in the first hours of ATRA treatment (peaking after 6h) and immediately down-regulated in TPA induced cells. Based on the obtained results, NB4 treated for 6h with TPA and 12h ATRA were included in C/EBPα PPI studies with PRISMA. Especially cells treated with TPA for prolonged periods of time (> 6h) were

going into apoptosis and started adhering to the cell culture dish, making preparation of nuclear extracts at later time points difficult. Since C/EBPα is of particular importance at early differentiation stages, including further differentiated cells was not expected to provide additional relevant PPI context.

For C/EBPα interactome studies, stable NB4 cell lines expressing inducible C/EBPα BioID fusion proteins were generated. Expression of C/EBPα-BioID in NB4 cells induced up-regulation of known C/EBPα target genes, like defensins and S100-A8, but did not lead to terminal differentiation of the cells as measured by CD11b surface marker expression. C/EBPα-BioID expression did not affect the response of NB4 cells to ATRA or TPA treatment. Why NB4 cells elude differentiation through C/EBPα in this context is not clear. One possible explanation is that the PML-RARA oncogene is repressing the expression of differentiation relevant genes and would need to be deactivated first. However, the absence of terminal differentiation by C/EBPα-BioID in these cells is not concerning in the context of this study. The up-regulation of C/EBPα target genes indicates that the fusion protein is active as a transcription factor and able to correctly dimerise and bind to DNA.

In a previous study that investigated the protein interaction landscape of C/EBPβ, PRISMA was performed with commercial nuclear protein extracts from HELA cells (Dittmar et al., 2019). For the C/EBPα PRISMA study presented in this thesis, nuclear extract from NB4 cells was used. Nuclear extraction was optimised and western blotting confirmed adequate purity of the nuclear fraction. Nuclear extracts generated with this protocol first published by Dignam et al., have been previously used for in vitro transcription assays (Dignam et al., 1983). This implies that major protein complexes and protein interactions are still intact in the extract and it is therefore expected that secondary interactions can also be detected by PRISMA.

4.2. Post-translational modifications of C/EBPα

C/EBP transcription factors contain numerous PTMs. The PhosphoSitePlus database (Hornbeck et al., 2012) contains 45 different side chain modifications for C/EBPβ, of which most are S,Y-phosphorylations. In addition, over 30 different arginine and lysine methylation sites have been characterised on C/EBPβ (Dittmar et al., 2019; Leutz et al., 2011). In contrast, only 15 different PTMs are annotated for C/EBPα in the PhosphoSitePlus database. Besides that, methylation of C/EBPα at R35, R156, R165 has been recently described (Li-ming et al., 2019). This difference in numbers of PTMs might be due to a lack of C/EBPα PTM annotation and not an actual difference in numbers. Published data from metabolic labelling with 3H-SAM and immune-affinity experiments with methyl-arginine/lysine specific antibodies suggested that the N-

termini of C/EBPα,β are extensively post-translationally modified by K,R-methylation (Kowenz-Leutz et al., 2010; Pless et al., 2008). Several K and R residues within the C/EBPα amino acid sequence are evolutionary conserved and possible methylation targets. Of particular interest in this regard is the CR1L region located in the N-terminus of P30-C/EBPα. CR1L (AA 131- 155) lies within a transactivating element of C/EBPα (TEIII; AA 126 - 200) and has been previously shown to interact with the SWI/SNF complex (Pedersen et al., 2001). A recent study suggested that clusters of arginine methylation within disordered region may provide a tunable protein interaction interfaces (Woodsmith et al., 2018) and such a regulatory mechanism is also conceivable for C/EBPα CR1L. In-house experiments have demonstrated that site-directed mutagenesis of three conserved arginine residues (R142, R149, R156) within CR1L alters transdifferentiation potential of C/EBPα (unpublished data).

Out of the three arginine residues in C/EBPα CR1L, only R156 monomethylation, as well as dimethylation, has been previously confirmed by mass spectrometry (Li-ming et al., 2019). Inspection of the CR1L amino acid sequence reveals there are no tryptic peptides spanning R142 shorter than 40 amino acids, which is above the upper limit for detection by traditional shotgun approaches. While digestion of protein samples with trypsin (cleaves C-terminal of arginine and lysine) and LysC (cleaves C-terminal of lysine) is most common, there are also other proteases available for shotgun proteomics. *In silico* digest of C/EBPα with AspN (cleaves N-terminal of aspartic acid and cysteic acid) and trypsin produces a peptide spanning R142 that can potentially be detected by mass spectrometry. Therefore, C/EBPα-BioID pull-downs were digested with different proteases (trypsin, LysC, AspN, chymotrypsin) and subjected to shotgun mass spectrometry. This approach detected known (R156, R35) as well as a novel (R12) C/EBPα arginine methylation/dimethylation site on C/EBPα. To increase sensitivity and facilitate the detection of R142me, a targeted parallel reaction monitoring (PRM) approach was employed. PRM measurements specifically monitoring a peptide with the sequence DGRmeLEPLEYER successfully confirmed C/EBPα methylation at R142. These results demonstrate that there are more arginine methylation sites on C/EBPα than the ones currently known. Certain PTM-sites might be elusive to large-scale detection methods because of technical issues, like ionisation and fragmentation problems (Sanders et al., 2007), and low abundance. Whether or not the CR1L arginine residue R147 is methylated remains to be confirmed.

PTMs located in the bZIP domain are mostly expected to modulate DNA binding or dimerisation. Modifications within transactivating or regulatory regions of C/EBPα, on the other hand, may directly influence PPIs or alter C/EBPα structure. In order to evaluate the influence of C/EBPα PTMs on protein interaction, 40 different PTMs were

included in protein interaction screening with PRISMA. Since PRISMA allows the high-throughput screening of peptide-protein interactions, all possible arginine- and lysine-methylations of C/EBPα outside of the bZIP domain were included. Methylation at R142 influenced over 100 protein interactions in PRISMA while no differential interactions were found with the R147 methylated peptide. Taking these results and time restrictions into consideration, no targeted assay was set up for the detection of R147 methylation by mass spectrometry.

4.3. C/EBPα protein interactions

Several studies have shown the importance of PPIs for C/EBPα functions and unraveling the C/EBPα interactome is anticipated to provide further biological insights (Johansen et al., 2001; Pedersen et al., 2001; Porse et al., 2001; Slomiany et al., 2000; Wang et al., 2001). C/EBPs contain intrinsically disordered regions in their N-termini that are implicated in many trans-regulatory processes. Protein interactions mediated by disordered regions are in general of low affinity but high specificity, with a certain degree of promiscuity, that allows dynamic regulation and rapid exchange of interaction partners (Wright and Dyson, 2015). The detection of these dynamic protein interactions is notoriously challenging. A previous study compared different immune purification strategies for C/EBPα and showed surprisingly low overlap (between 0 to 5%) between the individual purification strategies tested (Giambruno et al., 2013). This observed low overlap may be attributed to biochemical differences in the purification protocols and/or low reproducibility of antibody-based pull-downs of C/EBPα in general. PRISMA and BioID represent alternative and complementary methods for the detection of dynamic and SLiM-based protein interactions.

4.3.1. C/EBPα PRISMA screen

In PRISMA, peptides with and without PTMs are synthesised in an array format on a cellulose membrane support and screened for protein interactions (Dittmar et al., 2019; Meyer et al., 2018). The experimental success of the PRISMA approach may be attributed to the high local peptide concentration on the membrane (5 nmol peptide/ spot) which may counterbalance dissociation of weak interactors by molecular crowding and resulting rebinding effects (Ruthenburg et al., 2007). In the present study, 120 C/EBPα derived peptides, designed with a sequence overlap of seven amino acids, were screened for nuclear protein interactions with PRISMA. In total 40 different PTMs were integrated into the screen, including the newly identified C/EBPα arginine methylation sites at positions R12 and R142. PRISMA provided a detailed interaction

map of interacting proteins and presumably protein complexes across the C/EBPα amino acid sequence and modification sites. Hotspots for protein interaction correlated with conserved regions in C/EBPα. Many interacting proteins displayed multiple interactions with several conserved regions (CR2, CR3,4, and CR1L) located in the major transactivating regions.

Multivalent and redundant contacts between several sites on transcription factors and co-regulatory proteins have been observed before and may relate to dynamic and promiscuous aspects of the gene regulatory machinery (Brzovic et al., 2011; Clark et al., 2018; Currie et al., 2017; Vojnic et al., 2011). Fitting with this model, cooperativity of C/EBPα transactivating elements (TEs, see Figure 3) has previously been reported (Nerlov and Ziff, 1994, 1995). The authors of those studies reported that combinations of TEs functioned synergistically to recruit the TBP/TFIIB complex, an essential component of the RNA polymerase II basal transcription apparatus. By itself, the TEIII region, which corresponds to CR1L within the N-terminus of P30-C/EBPα, displayed no affinity to TBP/TFIIB. Concordantly, P30-C/EBPα that lacks the N-terminal transactivating TEI and TEII (CR2,3,4) is frequently described as a dominant inhibitor of the full-length P42 isoform. Here, PRISMA detected protein interaction hotspots not only in CR2, CR3 or 4 (TEI and TEII) that are unique to P42-C/EBPα but also in CR1L (TEIII). This indicates that the CR1L region in the P30-C/EBPα isoform shares interaction with a set of proteins and complexes that also interact with the P42-C/EBPα specific N-terminus. The data from PRISMA further suggest that PTMs may have an important function in orchestrating the dynamics of multivalent interactions of C/EBPα with major components of the transcriptional and epigenetic machinery.

Comparing the C/EBPα-PRISMA data with previously published data from a C/EBPβ-PRISMA screen (Dittmar et al., 2019) revealed striking similarities in the binding pattern of mediator complex components to homologous conserved regions in both C/EBPs, potentially reflecting the redundancy of C/EBPα/C/EBPβ function observed in many settings (Chen et al., 2002; Hirai et al., 2006; Jones et al., 2002). The mediator complex is essential for transcription and bridges between transcription factors binding to DNA with the core transcription machinery and RNA polymerase II (Soutourina, 2018). In addition to the mediator interaction sites CR2,3,4 contained in both C/EBPα and C/EBPβ, mediator components also bound to CR1L in C/EBPα. Other overlapping interactors were also found. However, due to significant differences in the two experimental PRISMA setups one cannot deduce functional differences of the two proteins from these datasets.

4.3.2. Validation of PRISMA with BioID

Proximity labelling in live cells, a technique suited for the detection of transient and dynamic interactions, was used to validate data derived from PRISMA. Although BioID was first introduced as proximity labelling to detect the spatial interactome of cellular structures, it has more recently also been employed to elucidate interactions of the transcriptional machinery (Kalkat et al., 2018; Kim et al., 2017). In BioID, proximal proteins are covalently modified with biotin and are subsequently enriched via highly stringent affinity purification protocols that remove contaminants and permit detection of low abundant interactors (Roux et al., 2012). BioID experiments in NB4 cells identified many known C/EBPα dimerisation partners, including other C/EBP and ATF transcription factors as the most enriched proteins. PRISMA and BioID datasets were integrated together with public databases to create a high confidence C/EBPα interaction map across C/EBPα regions and PTM sites. This network of linear interactors includes known and novel C/EBPα interactors that are highly connected by experimentally validated interactions deposited in the STRING database.

Known interactors that can now be mapped to the C/EBPα sequence include the histone acetyl-transferases P300/CBP and components of histone deacetylation and chromatin remodelling complexes. PRISMA mapped binding of P300/CBP to CR3/4 with residual binding in CR2 and CR1L. This binding pattern is in line with previous findings showing that a C/EBPα region spanning amino acids 55–108 (including CR3 and CR4) is sufficient, but not essential to mediate interaction with P300 and to induce adipogenesis (Erickson et al., 2001). SWI/SNF components (SMARCE1, SMARCA4, SMARCC1, SMARCC2, ACTL6A) also bound to the regions CR3/4 and CR1L. The CR1L region has previously been demonstrated to interact with the SWI/SNF complex (Müller et al., 2004). Components of the histone deacetylation complex NuRD (HDAC1,2) have previously been shown to interact with C/EBPα (Grebien et al., 2016). In PRISMA, NuRD subunits (HDAC1, HDAC2, GATAD2A, CHD4) interacted with C/EBPα peptides derived from CR3/4 and regions of low complexity located between CR5 and CR1L and the bZIP domain.

Novel C/EBPα interactors that were identified by both PRISMA and BioID include the transcription factors GABPA, GABPB1, FOXK1 and the cyclin-dependent kinases CDK9 and CDK13. GABPA and GABPB1 are part of the tetrameric transcription factor complex GABP that is required for myeloid differentiation. Disruption of Gabpa in mice is associated with a marked reduction in myeloid progenitor cells (Yang et al., 2011). In PRISMA the interaction of GABPA and GABPB1 mapped to CR1L and a region C-terminally adjacent to CR6. The novel C/EBPα interactors CDK9

and CDK13 regulate transcriptional elongation and mRNA maturation (Bacon and D'Orso, 2019; Greenleaf, 2019). Deregulation of CDK9 has been observed in several human malignancies and CDK9 inhibitors have recently shown encouraging clinical activity in newly diagnosed and relapsed AML (Wu et al., 2018). Like CDK13, CDK9 interacted with C/EBPα peptides corresponding to CR2, CR3,4 and CR1L. The transcription factor FOXK1 regulates glucose metabolism, differentiation and autophagy (Sakaguchi et al., 2019). Binding of FOXK1 was mapped to C/EBPα CR3,4 and CR1L.

As expected, a number of C/EBPα interactors were detected in only one of the two datasets. Some of the C/EBPα interactors may require more complex, simultaneous multi-site interactions or induced fit processes on the C/EBP structure, and were therefore missed by PRISMA. Discrepancies may further relate to the BioID preference for proximal interactions and the PRISMA preference for more abundant interactors as well as the detection of secondary interactors. The preference of PRISMA for more abundant interactors may be related to the experimental conditions, specifically the use of SILAC labeled extract and the mild washing conditions. Using SILAC adds complexity to the sample that results in a decreased number of peptide identifications while the mild washing steps lead to an increased background that can mask lower abundant proteins. Increasing the number of technical replicates for LFQ analysis and refraining from the use of SILAC in prospective PRISMA screens may facilitate the detection of low abundant proteins and increase statistical power.

Nevertheless, these datasets provide an extended C/EBPα interactome that may help to explain many functions of C/EBPα and provide the rationale for mutant design. Following this notion, the PRISMA detected R142 (CR1L) methylation-dependent interaction with the SWI/SNF complex subunit SMARCE1 was validated with BioID experiments. Proximity labelling using WT and methylation-mimicking C/EBPα mutant identified SMARCE1 together with other SWI/SNF components (ARID1A, ARID1B, ARID2) and subunits of the DREAM complex (LIN37, LIN9, MYBL2) as arginine methylation-specific interactors of C/EBPα.

4.3.3. Functional roles of conserved C/EBPα regions

In PRISMA, interaction hotspots were identified within short conserved C/EBPα regions. GO term enrichment of the validated interactors of each conserved region revealed that individual CRs are connected to distinct functional roles of C/EBPα. While the N-terminal region CR2 binds proteins connected to transcription, mRNA processing and cell cycle, the GO term chromatin remodelling was only enriched in CR1L and CR3,4. Although several proteins interacted with the regions CR5,6 and 7 only few GO terms were found enriched, suggesting functional heterogeneity. CR3,4 contained most of the significantly enriched GO terms and was also the only region that was enriched for the GO terms haematopoietic progenitor cell differentiation, histone acetylation and deacetylation. This indicates the importance of the core transactivating region CR3,4 to all P42-C/EBPα functions. The CR4 region of C/EBPα,β,δ,ε, previously also addressed as homology box B (Nerlov and Ziff, 1994), shares some homology with the HOB2 transactivating region in the bZIP transcription factors FOS and JUN (**Figure 24**). Deletion of the HOB2 region in FOS and JUN diminished transactivation potential of the two transcription factors (Sutherland et al., 1992). In lymphoid to myeloid transdifferentiation experiments with C/EBPβ, deletion of CR3 and CR4 almost completely abrogated reprogramming potential (Stoilova et al., 2013). Together with the data from PRISMA, these results suggest that C/EBP CR3,4 present interaction motifs for the recruitment of transactivating proteins and complexes, that are also contained in other transcription factors of the bZIP family.

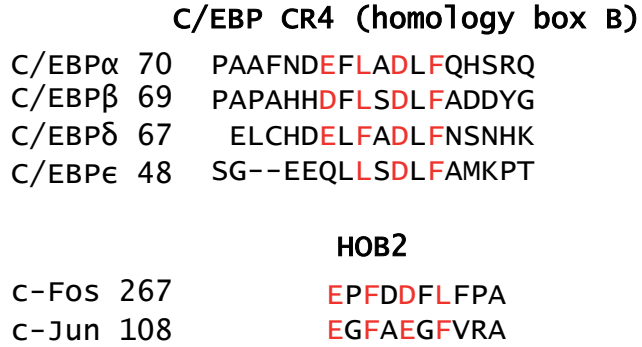


Figure 24: C/EBP region CR4 shows homology to the HOB2 region in FOS and JUN. Alignment of the conserved region CR4 of C/EBPα,β,δ,ε, also previously described as homology box B (Nerlov and Ziff, 1994), and the HOB2 region of the bZIP transcription factors FOS and JUN.

4.3.4. C/EBPα isoform-specific interactions

In order to further explore the C/EBPα isoform-specific interactome, proximity labelling experiments with P42- and P30-C/EBPα isoforms were performed. BioID confirmed interaction sites for transcription-associated proteins in P30 that were observed in PRISMA. While the majority of interactors were enriched in both pull-downs with similar efficiency, a subset of 71 and 9 proteins preferentially interacted with the P42 or P30 isoform respectively. P42-C/EBPα specific interactors identified with BioID showed a higher affinity for PRISMA peptides that correspond to the unique part of P42-C/EBPα. BioID interactors specific for P30 or shared between both C/EBPα isoforms displayed stronger binding to PRISMA peptides that were derived from P42/P30 shared regions. These results confirm that protein interaction mapping by PRISMA provides accurate information. However, BioID experiments identified several C/EBPα isoform-specific interactors that were not detected in PRISMA, possibly due to the preference of PRISMA for higher abundant interactors.

4.3.4.1. P42-C/EBPα specific interactors

The majority of proteins that specifically interacted with full-length C/EBPα are transcriptional regulators and activators. P42-C/EBPα specific interactors included several mediator subunits (MED1, MED12, MED15, MED17, MED20), transcription factors (EGR1, KLF5, ETV5, TFAP2A, FOXC1), the histone acetyl transferases CREBBP and EP300 as well as the nuclear receptor PPARG. The most prominent and novel P42-specific interactor in BioID was the early growth response protein 1 (EGR1), a zinc finger transcription factor involved in many biological processes. EGR1 was first described to be involved in growth and proliferation regulation, with RNA levels increasing during cardiac and neural cell differentiation (Sukhatme et al., 1988). EGR1 is very rapidly up-regulated after various stimuli including mitogen activation through ERK (Thiel and Cibelli, 2002), DNA damage (Quiñones et al., 2003) and drug treatment (Hu et al., 2010). Functions associated with EGR1 include differentiation, inhibition of cell growth and proapoptotic functions but also growth-promoting functions, depending on cell type and stimulus (Chen et al., 2019; Group, 2001; Thiel and Cibelli, 2002; Yu et al., 2007). In myeloid cells, EGR1 acts as a positive regulator of differentiation. EGR1 is essential for and differentiation along the macrophage lineage (Nguyen et al., 1993) and in mice, Egr1 and Egr2 are positive regulators of macrophage differentiation under the instruction of the myeloid transcription factor PU.1 (Laslo et al., 2006). Expression of EGR1 in the murine myeloid cell line 32Dcl3 induces macrophage differentiation at the expense of granulocytes and erythrocytes (Krishnaraju et al., 2001) and in NB4

cells, EGR1 is up-regulated during the first hours of TPA treatment. In hepatocytes, EGR1 expression is stimulated by glucagon (Shen et al., 2015) and the authors of the study found that EGR1 bound to the C/EBPα promoter. Another study in a liver cancer cell line reported that C/EBPβ but not C/EBPα interacts with EGR1 at the LDLR promoter (Zhang et al., 2003). In the BioID experiments in NB4 cells that are described here, EGR1 was detected as an interactor of P42 but not P30, indicating that the protein interacts with the N-terminal region of C/EBPα. EGR1 was not detected in PRISMA, potentially attributed to the low abundance of the protein (ranked by intensity, EGR was on position 6221 of 6450 detected proteins in the PRISMA input material). Based on the available data, it is tempting to speculate that C/EBPα P42 and EGR1 interact in myeloid cells and coordinate myeloid differentiation. A study in mice liver regeneration has previously suggested that EGR1 can either directly interact with DNA or indirectly through the interaction with CEBPs (Gallardo et al., 2016).

Another interactor of C/EBPα P42 that did not interact with the P30 isoform in BioID experiments was the nuclear receptor and master regulator of adipogenesis PPARγ. During adipocyte maturation, PPARγ and C/EBPα coordinately orchestrate the adipogenic gene program (Lefterova et al., 2008). Apart from its pivotal role in adipocytes PPARγ is also expressed in other cells, and while it is dispensable for macrophage differentiation, PPARγ is essential for the establishment of an anti-inflammatory phenotype in adipose tissue macrophages (Lefterova et al., 2014). Results from ChIP profiling revealed that the genomic sites occupied by PPARγ in macrophages overlap with binding sites of the myeloid transcription factor PU.1 and C/EBPβ (Lefterova et al., 2010; Pott et al., 2012). Microarray data analysis of NB4 cells employed in this study hints that PPARγ functionally interacts with P42-, but not P30-C/EBPα to activate lineage specific gene programs in myeloid cells.

4.3.4.2. P30-C/EBPα specific interactors

Since the entirety of P30 is contained in the P42 isoform, it is counterintuitive that P30 specifically interacts with different proteins compared to P42. It is possible that different biophysical properties, like PTM status or conformation, result in differential protein interaction patterns. BioID detected 9 proteins that specifically interacted with P30-C/EBPα in NB4 cells (EHMT2, GATA1, BLM, TFAP4, CENPC1, ZNF362, AHDC1, BCL11A, EBF3BF1). Except for BLM and CENPC1, most of P30-specific interactors were also enriched in P42 BioID pull-downs, albeit with less efficiency compared to P30. This indicates that the majority of P30 interactors also interact with P42 and the interaction might also be regulated by isoform abundance in the cell. P30-C/EBPα acts

as an oncogene in AML and isoform-specific interactors may provide new insight for potential therapeutic intervention.

The P30-specific interactor GATA1 is a master regulator of erythroid development. In animal models, conditional GATA1 knockout causes X-chromosome-linked anemic or bleeding disease due to defects in the formation of red blood cells and platelets (Fujiwara et al., 1996). In murine erythroleukaemia cells, Gata1 repressed Cebpa expression by negatively regulating the myeloid/lymphoid transcription factor Pu.1 (Burda et al., 2009). Here, microarray analysis of C/EBPα isoform expressing NB4 cells revealed that a GATA1 signature correlated with P30- but not P42-C/EBPα expression. This provides further evidence that GATA1 either physically interacts with or resides in very close proximity to P30-C/EBPα.

The methyltransferase EHMT2, also known as G9a, specifically interacted with P30-C/EBPα. EHMT2 catalyses the mono- and dimethylation of histone H3 at lysine 9 and 27 and deregulated expression is implicated in human malignancies. In AML mouse models, loss of EHMT2 expression significantly delayed disease progression and reduced leukaemia stem cell frequency (Lehnertz et al., 2014). Initially identified as an activator, the P30 interactor TFAP4 has been found to cooperate with RUNX proteins in gene silencing in the haematopoietic system (Egawa and Littman, 2011). In neuroblastoma cells, TFAP4 is a target of MYCN and down-regulation of TFAP4 expression led to inhibition of cell proliferation and migration (Boboila et al., 2018). Knockout of the TFPA4 gene was also shown to slow growth in 9 out of 14 AML cell lines (Meyers et al., 2017; Tsherniak et al., 2017).

A previous study investigating the C/EBPα isoform-specific interactome performed in the murine lymphoblastic cell line FDCP-1 reported large qualitative differences in the interactomes C/EBPα isoforms. Although there are some overlaps to the interactome presented here, the BioID data does not support this finding and did not confirm any isoform-specific interactor reported by Grebien et al. In total, 10 C/EBPα interactors detected by BioID were also contained in the Grebien et al. data, including the histone deacetylases HDAC1,2 and the MLL complex subunit WDR5. The differential evaluation of C/EBPα isoform-specific interactomes here and by Grebien et al. is probably due to differences in the experimental setup and data analysis workflow. The approach by Grebien et al. is based on a label-free analysis of a single biological replicate without direct quantitative or statistical comparison. Additionally, the negative control chosen by the authors – mock infected cells that do not contain the affinity tag – does not allow to distinguish true C/EBPα interactors from background proteins that bind to the tag alone. On the other hand, the data from BioID are based on a quantitative comparison of four biological replicates and rigorous statistical testing

against background controls. BioID data are in line with findings from PRISMA and provides strong evidence that the interactomes of P42/P30-C/EBPα largely overlap with quantitative differences in the affinity of a subset of interactors. P30 might therefore not act as a dominant-suppressor of P42-C/EBPα, or an entirely different transcription factor, but rather as a transcriptionally weaker derivative that may moderate several functions of P42-C/EBPα. In support of this interpretation, a previous study with P42-C/EBPα deficient mice engineered to express P30-C/EBPα from the Cebpa locus showed rescue of lethality, facilitated myeloid progenitor commitment, increased proliferation and experimental confirmation of an AML oncoprotein with complete penetrance (Bereshchenko et al., 2009; Kirstetter et al., 2008).

4.4. Conclusion and outlook

C/EBPα is a lineage specific transcription factor characterised by high intrinsic disorder and numerous PTMs decorating its sequence. Data from PRISMA and BioID experiments were integrated to provide a comprehensive C/EBPα interactome mapped across C/EBPα conserved regions and PTM sites. Using myeloid cells as a model system facilitated the detection of novel myeloid-specific C/EBPα interactors. The C/EBPα interaction map presented in this thesis may serve as a resource for many further studies exploring the functionality and biological importance of individual C/EBPα regions and PTMs. In the future, using the newest generation of proximity biotin ligases (TurboID) would also allow probing for interactors with a short half-life and elucidating the dynamics of protein interactions during differentiation.

Analogously to beads on a string, regions of high intrinsic disorder and low sequence complexity interconnect conserved C/EBP regions. The data presented here support the hypothesis that the interaction of C/EBPα with components of the transcriptional and epigenetic machinery is coordinated by multivalent interactions with short interaction motifs in C/EBPα conserved regions. How these interactions are assembled in a three-dimensional architecture is an intriguing question that will require further experiments.

Similarities of C/EBPα PRISMA with C/EBPβ protein interaction profiles from a previous PRISMA screen indicate that certain protein interactions are conserved among C/EBP factors. Beyond the functional analysis of C/EBPs, this study suggests that the integration of PRISMA and BioID is a favourable strategy to explore the linear and PTM dependent interactome of a vast number of intrinsically disordered proteins involved in cell signalling and gene regulation.

5. References

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198–207.

Amoutzias, G.D., Veron, A.S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S.G., and Robertson, D.L. (2007). One billion years of bZIP transcription factor evolution: Conservation and change in dimerisation and DNA-binding site specificity. Mol. Biol. Evol. *24*, 827–835.

Andreasson, C., and Ljungdahl, P.O. (2004). The N-Terminal Regulatory Domain of Stp1p Is Modular and, Fused to an Artificial Transcription Factor, Confers Full Ssy1p-Ptr3p-Ssy5p Sensor Control. Mol. Cell. Biol. *24*, 7503–7513.

Avellino, R., and Delwel, R. (2017). Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation. Blood *129*, 2083–2091.

Babu, M.M. (2009). The rules of disorder or why disorder rules. Prog. Biophys. Mol. Biol. *99*, 94–103.

Bacon, C.W., and D'Orso, I. (2019). CDK9: a signalling hub for transcriptional control. Transcription *10*, 57–75.

Bagert, J.D., Xie, Y.J., Sweredoski, M.J., Qi, Y., Hess, S., Schuman, E.M., and Tirrell, D.A. (2014). Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labelling by Amino Acids in Cell Culture. Mol. Cell. Proteomics *13*, 1352–1358.

Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. Anal. Bioanal. Chem. *404*, 939–965.

Barbar, E., and Nyarko, A. (2015). Polybivalency and disordered proteins in ordering macromolecular assemblies. Semin. Cell Dev. Biol. *37*, 20–25.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Månsson, R., Luc, S., Grover, A., Jacobsen, S.E.W., Bryder, D., and Nerlov, C. (2009). Haematopoietic Stem Cell Expansion Precedes the Generation of Committed Myeloid Leukemia-Initiating Cells in C/EBPα Mutant AML. Cancer Cell *16*, 390–400.

Biggar, K.K., and Li, S.S.-C. (2015). Non-histone protein methylation as a regulator of cellular signalling and function. Nat. Rev. Mol. Cell Biol. *16*, 5–17.

Birkenmeier, E.H., Gwynn, B., Howard, S., Jerry, J., Gordon, J.I., Landschulz, W.H., and McKnight, S.L. (1989). Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev. *3*, 1146–1156.

Boboila, S., Lopez, G., Yu, J., Banerjee, D., Kadenhe-chiweshe, A., Connolly, E.P., Kandel, J.J., Rajbhandari, P., Silva, J.M., Califano, A., et al. (2018). Transcription factor activating protein 4 is synthetically lethal and a master regulator of MYCN- ampli fi ed neuroblastoma. Oncogene 5451–5465.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. *28*, 420–435.

Bozorgzadeh, M.H., Morgan, R.P., Beynon, J.H.,(1978) Application of Mass-analysed Ion Kinetic Energy Spectrometry (MIKES) to the Determination of the Structures of Unknown Compounds, Analyst, 103, 613-622

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labelling in living cells and organisms with TurboID. Nat. Biotechnol. *36*, 880–887.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729–736.

Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., et al. (2011). The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Mol. Cell *44*, 942–953.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. Nucleic Acids Res. *45*, D369–D379.

Chen, L., Wang, S., Zhou, Y., Wu, X., Entin, I., Epstein, J., Yaccoby, S., Xiong, W., Barlogie, B., Jr, J.D.S., et al. (2019). Identification of early growth response protein 1 (EGR-1) as a novel target for JUN -induced apoptosis in multiple myeloma. *115*, 61–71.

Chen, S.-S., Chen, J.-F., Johnson, P.F., Muppala, V., and Lee, Y.-H. (2002). C/EBPbeta , When Expressed from the C/ebpalpha Gene Locus, Can Functionally Replace C/EBPalpha in Liver but Not in Adipose Tissue. Mol. Cell. Biol. *20*, 7292–7299.

Chinetti, G., Griglio, S., Antonucci, M., Torra, I.P., Delerive, P., Majd, Z., Fruchart, J.C., Chapman, J., Najib, J., and Staels, B. (1998). Activation of proliferator-activated receptors α and γ induces apoptosis of human monocyte-derived macrophages. J. Biol. Chem. *273*, 25573–25580.

Christensen, D.E., and Klevit, R.E. (2009). Dynamic interactions of proteins in complex networks : identifying the complete set of interacting E2s for functional investigation of E3-dependent protein ubiquitination. FEBS J. *276*, 5381–5389.

Cirilli, M., Bereshchenko, O., Ermakova, O., and Nerlov, C. (2017). Insights into specificity, redundancy and new cellular functions of C/EBPa and C/EBPb transcription factors through interactome network analysis. Biochim. Biophys. Acta - Gen. Subj. *1861*, 467–476.

Cirovic, B., Schönheit, J., Kowenz-leutz, E., Ivanovska, J., Klement, C., Pronina, N., Bégay, V., and Leutz, A. (2017). C/EBP-Induced Transdifferentiation Reveals Granulocyte-Macrophage Precursor-like Plasticity of B Cells. Stem Cell Reports *8*, 346–359.

Clark, S., Myers, J.B., King, A., Fiala, R., Novacek, J., Pearce, G., Heierhorst, J., Reichow, S.L., and Barbar, E.J. (2018). Erratum: Multivalency regulates activity in an intrinsically disordered transcription factor (eLife (2018) 7 PII: e40684). Elife *7*, 1–28.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526.

Cronan, J.E. (2005). Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. J. Nutr. Biochem. *16*, 416–418.

Currie, S.L., Doane, J.J., Evans, K.S., Bhachech, N., Madison, B.J., Lau, D.K.W., McIntosh, L.P., Skalicky, J.J., Clark, K.A., and Graves, B.J. (2017). ETV4 and AP1 Transcription Factors Form Multivalent Interactions with three Sites on the MED25 Activator-Interacting Domain. J. Mol. Biol. *429*, 2975–2995.

Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., and Spiegelman, B. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately down-regulated in human diabetes. Nat. Genet. *34*, 267–273.

Darling, D., Hughes, C., Galea-Lauri, J., Gäken, J., Trayner, I.D., Kuiper, M., and Farzaneh, F. (2000). Low-speed centrifugation of retroviral vectors absorbed to a particulate substrate: A highly effective means of enhancing retroviral titre. Gene Ther. *7*, 914–923.

Daughdrill, G.W., Chadsey, M.S., Karlinsey, J.E., Hughes, K.T., and Dahlquist, F.W. (1997). The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, σ28. Nat. Struct. Biol. *4*, 285–291.

Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., and Gibson, T.J. (2012). Attributes of short linear motifs. Mol. BioSyst. *8*, 268–281.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *1*, 1475–1489.

Dittmar, G., Hernandez, D.P., Kowenz-Leutz, E., Kirchner, M., Kahlert, G., Wesolowski, R., Baum, K., Knoblich, M., Hofstätter, M., Muller, A., et al. (2019). PRISMA: Protein Interaction Screen on Peptide Matrix Reveals Interaction Footprints and Modifications- Dependent Interactome of Intrinsically Disordered C/EBPβ. IScience *13*, 351–370.

Dunham, W.H., Mullin, M., and Gingras, A.C. (2012). Affinity-purification coupled to mass spectrometry: Basic principles and strategies. Proteomics *12*, 1576–1590.

Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J.E. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac. Symp. Biocomput. 473–484.

Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. Genome Inform. Ser. Workshop Genome Inform. *11*, 161–171.

Dyson, H.J., and Wright, P.E. (2004). Unfolded Proteins and Protein Folding Studied by NMR. Chem. Rev. *104*, 3607–3622.

Dyson, H.J., and Wright, P.E. (2005). INTRINSICALLY UNSTRUCTURED PROTEINS AND THEIR FUNCTIONS. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Dyson, H.J., and Wright, P.E. (2016). Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding Protein (CBP) and p300. J. Biol. Chem. *291*, 6714–6722.

Egawa, T., and Littman, D.R. (2011). Transcription factor AP4 modulates reversible and epigenetic silencing of the Cd4 gene. Proc. Natl. Acad. Sci. U. S. A. *108*, 14873–14878.

Erickson, R.L., Hemati, N., Ross, S.E., and MacDougald, O.A. (2001). p300 Coactivates the Adipogenic Transcription Factor CCAAT/ Enhancer-binding Protein alpha*. J. Biol. Chem. *276*, 16348–16355.

Fasan, A., Haferlach, C., Alpermann, T., Jeromin, S., Grossmann, V., Eder, C., Weissmann, S., Dicker, F., Kohlmann, A., Schindela, S., et al. (2014). The role of different genetic subtypes of CEBPA mutated AML. Leukemia *28*, 794–803.

Ferrari-Amorotti, G., Mariani, S.A., Novi, C., Cattelani, S., Pecorari, L., Corradini, F., Soliera, A.R., Manzotti, G., Fragliasso, V., Zhang, Y., et al. (2010). The biological effects of C/EBPα in K562 cells depend on the potency of the N-terminal regulatory region, not on specificity of the DNA binding domain. J. Biol. Chem. *285*, 30837–30850.

Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006). A practical overview of protein disorder prediction methods. Proteins *65*, 1–14.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. Berichte Der Dtsch. Chem. Gesellschaft *27*, 2985–2993.

Forman-kay, J.D., and Mittag, T. (2013). From Sequence and Forces to Structure , Function , and Evolution of Intrinsically Disordered Proteins. Structure *21*, 1492–1499.

Forsström, B., Axnäs, B.B., Stengele, K., Bühler, J., Albert, T.J., Richmond, T.A., Hu, F.J., Nilsson, P., Hudson, E.P., Rockberg, J., et al. (2014). Proteome-wide Epitope Mapping of Antibodies Using Ultra-dense Peptide Arrays. Mol. Cell. Proteomics *13*, 1585–1597.

Frank, R. (1992). Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. Tetrahedron *48*, 9217–9232.

Frank, R. (2002). The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications. J. Immunol. Methods *267*, 13–26.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc. Natl. Acad. Sci. U. S. A. *93*, 12355–12358.

Garner, Cannon, Romero, Obradovic, and Dunker (1998). Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. Genome Inform. Ser. Workshop Genome Inform. *9*, 201–213.

Geletu, M., Balkhi, M.Y., Peer Zada, A.A., Christopeit, M., Pulikkan, J.A., Trivedi, A.K., Tenen, D.G., and Behre, G. (2007). Target proteins of C/EBP p30 in AML: C/EBP p30 enhances sumoylation of C/EBP p42 via up-regulation of Ubc9. Blood *110*, 3301–3309.

Giambruno, R., Grebien, F., Stukalov, A., Knoll, C., Planyavsky, M., Rudashevskaya, E.L., Colinge, J., Superti-furga, G., and Bennett, K.L. (2013). Affinity Purification Strategies for Proteomic Analysis of Transcription Factor Complexes. J. Proteome Res. *12*, 4018–4027.

Graves, B.J., Johnson, P.F., and Mcknight, S.L. (1986). Homologous Recognition of a Promoter Domain Common to the MSV LTR and the HSV tk Gene. Cell *44*, 565–576.

Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Barsyte-lovejoy, D., Li, F., et al. (2016). Pharmacological targeting of the Wdr5-MLL interaction in C / EBPα N-terminal leukemia. Nat. Chem. Biol. *11*, 571–578.

Green, N.. (1963). AVIDIN 1. THE USE OF [14C]BIOTIN FOR KINETIC STUDIES AND FOR ASSAY. Biochem. J. *89*, 585–591.

Greenleaf, A.L. (2019). Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. Transcription *10*, 91–110.

Group, N.P. (2001). Impaired prostate tumorigenesis in Egr1-deficient mice. 101–107.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput. Biol. *2*, 0890–0901.
He, B., Wang, K., Liu, Y., Xue, B., Uversky, V.N., and Dunker, A.K. (2009). Predicting intrinsic disorder in proteins: An overview. Cell Res. *19*, 929–949.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. Cell *163*, 712–723.

Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S.I., Imanishi, J., Akashi, K., and Tenen, D.G. (2006). C/EBPβ is required for “emergency” granulopoiesis. Nat. Immunol. *7*, 732–739.

Hornbeck, P. V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, 261–270.

Hu, C.T., Chang, T.Y., Cheng, C.C., Liu, C.S., Wu, J.R., Li, M.C., and Wu, W.S. (2010). Snail associates with EGR-1 and SP-1 to upregulate transcriptional activation of p15INK4b. FEBS J. *277*, 1202–1218.

Huafeng, X., Min, Y., Feng, R., and Graf, T. (2004). Stepwise Reprogramming of B Cells into Macrophages. Cell *117*, 663–676.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz J., Giordani A.B. (1981), Sequence Analysis of Polypeptides by Collision Activated Dissociation on a Triple Quadrupole Mass Spectrometer, Biological Mass Spectrometry, 8(9), 397-408

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., and Dunker, A.K. (2002). Intrinsic Disorder in Cell-signalling and Cancer-associated Proteins. J. Mol. Biol. *2836*, 573–584.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., Connor, T.R.O., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. *32*, 1037–1049.

Ivarsson, Y., and Jemth, P. (2019). Affinity and specificity of motif-based protein – protein interactions. Curr. Opin. Struct. Biol. *54*, 26–33.

Johansen, L.M., Iwama, A., Lodie, T.A., Sasaki, K., Felsher, D.W., Golub, T.R., and Tenen, D.G. (2001). c-Myc Is a Critical Target for C / EBPa in Granulopoiesis. Mol. Cell. Biol. *21*, 3789–3806.

Johnson, P.F., Landschulz, W.H., Graves, B.J., and Mcknight, S.L. (1987). Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Dev. *1*, 133–146.

Jones, L.C., Lin, M.L., Chen, S.S., Krug, U., Hofmann, W.K., Lee, S., Lee, Y.H., and Phillip Koeffler, H. (2002). Expression of C/EBPβ from the C/ebpα gene locus is sufficient for normal haematopoiesis in vivo. Blood *99*, 2032–2036.

Kalkat, M., Resetca, D., Lourenco, C., Chan, P.K., Wei, Y., Shiah, Y.J., Vitkin, N., Tong, Y., Sunnerhagen, M., Done, S.J., et al. (2018). MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. Mol. Cell *72*, 836-848.e7.

Katz, C., Levy-beladev, L., Rotem-bamberger, S., Rito, T., Rüdiger, S.G.D., and Friedler, A. (2011). Studying protein–protein interactions using peptide arrays. Chem.Soc.Rev. *44*, 2131–2145.

Keilhauer, E.C., Hein, M.Y., and Mann, M. (2014). Accurate protein complex retrieval by affinity enrichment MS rather than affinity purification MS. Mol. Cell. Proteomics. 14(1), 120-135

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature *181*, 662–666.

Kim, B.R., Coyaud, E., Laurent, E.M.N., St-Germain, J., Van De Laar, E., Tsao, M.S., Raught, B., and Moghal, N. (2017). Identification of the SOX2 interactome by BioID reveals EP300 as a mediator of SOX2-dependent squamous differentiation and lung squamous cell carcinoma growth. Mol. Cell. Proteomics *16*, 1864–1888.

Kim, D.I., Jensen, S.C., Noble, K.A., KC, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for BioID proximity labelling. Mol. Biol. Cell *27*, 1188–1196.

Kim, P.M., Sboner, A., Xia, Y., and Gerstein, M. (2008). The role of disorder in interaction networks: A structural analysis. Mol. Syst. Biol. *4*, 179.

Kirstetter, P., Schuster, M.B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Mönch, K., Månsson, R., Pedersen, T.Å., Pabst, T., et al. (2008). Modeling of C/EBPα Mutant Acute Myeloid Leukemia Reveals a Common Expression Signature of Committed Myeloid Leukemia-Initiating Cells. Cancer Cell *13*, 299–310.

Kitov, P.I., and Bundle, D.R. (2003). On the Nature of the Multivalency Effect: A Thermodynamic Model. J. Am. Chem. Soc. *125*, 16271–16284.

Kowenz-Leutz, E., Pless, O., Dittmar, G., Knoblich, M., and Leutz, A. (2010). Crosstalk between C/EBPB phosphorylation, arginine methylation, and SWI/SNF/Mediator implies an indexing transcription factor code. EMBO J. *29*, 1105–1115.

Krishnaraju, K., Hoffman, B., and Liebermann, D.A. (2001). Early growth response gene 1 stimulates development of haematopoietic progenitor cells along the macrophage lineage at the expense of the granulocyte and erythroid lineages. Blood *97*, 1298–1305.

Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I., and Wright, P.E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc. Natl. Acad. Sci. U. S. A. *93*, 11504–11509.

Laganowsky, A., Reading, E., Hopper, J.T.S., and Robinson, C. V (2014). Mass Spectrometry of Intact Membrane Protein Complexes. Nat. Protoc. *8*, 639–651.

Lange, S., Sylvester, M., Schümann, M., Freund, C., and Krause, E. (2010). Identification of phosphorylation-dependent interaction partners of the adapter protein ADAP using quantitative mass spectrometry: SILAC vs 18O-labelling. J. Proteome Res. *9*, 4113–4122.

Lanotte, M., Martin-Thouvenin, V., Najman, S., Balerini, P., Valensi, F., and Berger, R. (1991). NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). Blood *77*, 1080–1086.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006). Multilineage Transcriptional Priming and Determination of Alternate Haematopoietic Cell Fates. Cell *126*, 755–766.

Leblanc, S.J., Kulkarni, P., and Weninger, K.R. (2018). Single molecule FRET: A powerful tool to study intrinsically disordered proteins. Biomolecules *8*.

Lefterova, M.I., Zhang, Y., Steger, D.J., Schupp, M., Schug, J., Cristancho, A., Feng, D., Zhuo, D., Stoeckert, C.J., Liu, X.S., et al. (2008). PPARγ and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. Genes Dev. *22*, 2941–2952.

Lefterova, M.I., Steger, D.J., Zhuo, D., Qatanani, M., Mullican, S.E., Tuteja, G., Manduchi, E., Grant, G.R., and Lazar, M.A. (2010). Cell-Specific Determinants of Peroxisome Proliferator-Activated Receptor gamma Function in Adipocytes and Macrophages. Mol. Cell. Biol. *30*, 2078–2089.

Lefterova, M.I., Haakonsson, A.K., Lazar, M.A., and Mandrup, S. (2014). PPARg and the global map of adipogenesis and beyond. Trends Endocrinol. Metab. *25*.

Lehnertz, B., Pabst, C., Su, L., Miller, M., Liu, F., Yi, L., Zhang, R., Krosl, J., Yung, E., Kirschner, J., et al. (2014). The methyltransferase G9a regulates HoxA9-dependent transcription in AML. Genes Dev. *28*, 317–327.

Lekstrom-Himes, J., and Xanthopoulos, K.G. (1998). Biological role of the CCAAT/enhancer-binding protein family of transcription factors. J Biol Chem *273*, 28545–28548.

Leutz, A., Pless, O., Lappe, M., Dittmar, G., and Kowenz-, E. (2011). Crosstalk between phosphorylation and multi-site arginine/lysine methylation in C/EBPs. Transcription *2*, 3–8.

Li-ming, L., Wen-zheng, S., Xue-zhe, F., Ya-li, X., Mo-bing, C., and Ye, Z. (2019). Methylation of C/EBPα by PRMT1 inhibits its tumor suppressive function in breast cancer. Cancer Reaseatch.

Li, M., and Song, J. (2007). The N- and C-termini of the human Nogo molecules are intrinsically unstructured: Bioinformatics, CD, NMR characterization, and functional implications. Proteins Struct. Funct. Bioinforma. *68*, 100–108.

Liang, K., Gao, X., Gilmore, J.M., Florens, L., Washburn, M.P., Smith, E., and Shilatifard, A. (2015). Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. Mol. Cell. Biol. *35*, 928–938.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Lin, L.I., Chen, C.Y., Lin, D.T., Tsay, W., Tang, J.L., Yeh, Y.C., Shen, H.L., Su, F.H., Yao, M., Huang, S.Y., et al. (2005). Characterization of CEBPA mutations in acute myeloid leukemia: Most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin. Cancer Res. *11*, 1372–1379.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, a K. (2006). Intrinsic disorder in transcription factors.

Luchinat, E., and Banci, L. (2016). A unique tool for cellular structural biology: In-cell NMR. J. Biol. Chem. *291*, 3776–3784.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

Madsen, M.S., Siersbaek, R., Boergesen, M., Nielsen, R., and Mandrup, S. (2014). Peroxisome Proliferator-Activated Receptor and C/EBP Synergistically Activate Key Metabolic Adipocyte Genes by Assisted Loading. Mol. Cell. Biol. *34*, 939–954.

Majello, B., De Luca, P., and Lania, L. (1997). Sp3 is a bifunctional transcription regulator with modular independent activation and repression domains. J. Biol. Chem. *272*, 4021–4026.

Martell, J.D., Deerinck, T.J., Sancak, Y., Poulos, T.L., Mootha, V.K., Sosinsky, G.E., Ellisman, M.H., and Ting, A.Y. (2012). Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat. Biotechnol. *30*, 1143–1148.

McKnight, S.L. (1991). Molecular Zippers in Gene Regulation. Sci. Am. *264*, 54–64.

Metallo, S.J. (2011). Intrinsically disordered proteins are potential drug targets. Curr. Opin. Chem. Biol. *14*, 481–488.

Meyer, K., and Selbach, M. (2015). Quantitative affinity purification mass spectrometry: A versatile technology to study protein-protein interactions. Front. Genet. *6*, 1–7.

Meyer, K., Kirchner, M., Uyar, B., Cheng, J.Y., Russo, G., Hernandez-Miranda, L.R., Szymborska, A., Zauber, H., Rudolph, I.M., Willnow, T.E., et al. (2018). Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. Cell *175*, 239-253.e17.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N. V, Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Miotto, B., and Struhl, K. (2006). Differential Gene Regulation by Selective Association of Transcriptional Coactivators and bZIP DNA-Binding Domains. Mol. Cell. Biol. *26*, 5969–5982.

Mollica, L., Bessa, L.M., Hanoulle, X., Jensen, M.R., Blackledge, M., and Schneider, R. (2016). Binding Mechanisms of Intrinsically Disordered Proteins : Theory , Simulation , and Experiment. Front. Mol. Biosci. *3*, 1–18.

Müller, C., Calkhoven, C.F., Sha, X., and Leutz, A. (2004). The CCAAT Enhancer-binding Protein α (C/EBPα) Requires a SWI/SNF Complex for Proliferation Arrest. J. Biol. Chem. *279*, 7353–7358.

Nakajima, H., Watanabe, N., Shibata, F., Kitamura, T., Ikeda, Y., and Handa, M. (2006). N-terminal region of CCAAT/enhancer-binding protein epsilon is critical for cell cycle arrest, apoptosis, and functional maturation during myeloid differentiation. J. Biol. Chem. *281*, 14494–14502.

Nerlov, C., and Ziff, E.B. (1994). Three levels of functional interaction determine the activity of CCAAT/enhancer binding protein-α on the serum albumin promoter. Genes Dev. *8*, 350–362.

Nerlov, C., and Ziff, E.B. (1995). CCAAT / enhancer binding protein-a amino acid motifs with dual TBP and TFIIB binding ability co-operate to activate transcription in both yeast and mammalian cells. EMBO J. *14*, 4318–4328.

Nguyen, H.Q., Hoffman-liebermann, B., and Liebermann, D.A. (1993). The zinc finger transcription factor Egr-1 is essential for and restricts differentiation along the macrophage lineage. Cell *72*, 197–209.

Nielsen, J.T., and Mulder, F.A.A. (2019). Quality and bias of protein disorder predictors. Sci. Rep. *9*, 1–11.

Okada, H., Uezu, A., Soderblom, E.J., Moseley, M.A., Gertler, F.B., and Soderling, S.H. (2012). Peptide array X-linking (PAX): A new peptide-protein identification approach. PLoS One *7*, 1–10.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable Isotope Labelling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol. Cell. Proteomics *1*, 376–386.

Pabst, T., and Mueller, B.U. (2009). Complexity of CEBPA dysregulation in human acute myeloid leukemia. Clin. Cancer Res. *15*, 5303–5307.

Pedersen, T.Å., Kowenz-leutz, E., Leutz, A., and Nerlov, C. (2001). Cooperation between C / EBPalpha TBP / TFIIB and SWI / SNF recruiting domains is required for adipocyte differentiation. Genes Dev. *15*, 3208–3216.

Pless, O., Kowenz-Leutz, E., Knoblich, M., Lausen, J., Beyermann, M., Walsh, M.J., and Leutz, A. (2008). G9a-mediated lysine methylation alters the function of CCAAT/Enhancer- binding protein-β. J. Biol. Chem. *283*, 26357–26363.

Porse, B.T., Pedersen, T.Å., Xu, X., Lindberg, B., Wewer, U.M., Friis-hansen, L., and Nerlov, C. (2001). E2F Repression by C / EBPa Is Required for Adipogenesis and Granulopoiesis In Vivo. Cell *107*, 247–258.

Pott, S., Kamrani, N.K., Bourque, G., Pettersson, S., and Liu, E.T. (2012). PPARG Binding Landscapes in Macrophages Suggest a Genome-Wide Contribution of PU . 1 to Divergent PPARG Binding in Human and Mouse. PLoS One *7*.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: A general procedure of protein complex purification. Methods *24*, 218–229.

Quiñones, A., Dobberstein, K.U., and Rainov, N.G. (2003). The egr-1 gene is induced by DNA-damaging agents and non-genotoxic drugs in both normal and neoplastic human cells. Life Sci. *72*, 2975–2992.

Radomska, H.S., Huettner, C.S., Zhang, P., Cheng, T., Scadden, D.T., and Tenen, D.G. (1998). CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. Mol. Cell. Biol. *18*, 4301–4314.

Ramji, D.P., and Foka, P. (2002). CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem. J. *365*, 561–575.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionisation, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Reineke, U., and Sabat, R. (2008). Antibody Epitope Mapping Using SPOT ™ Peptide Arrays. Methods Mol. Biol. *524*, 145–167.

Rhee, H., Zou, P., Udeshi, N.D., Martell, J.D., Vamsi, K.M., Carr, S.A., and Ting, A.Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially- Restricted Enzymatic Tagging. Science. *339*, 1328–1331.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Rőszer, T., Menéndez-Gutiérrez, M.P., Lefterova, M.I., Alameda, D., Núñez, V., Lazar, M.A., Fischer, T., and Ricote, M. (2011). Autoimmune Kidney Disease and Impaired Engulfment of Apoptotic Cells in Mice with Macrophage Peroxisome Proliferator-Activated Receptor γ or Retinoid X Receptor α Deficiency. J. Immunol. *186*, 621–631.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810.

Roux, K.J., Kim, D.I., Burke, B., and May, D.G. (2018). BioID: A Screen for Protein-Protein Interactions. Curr. Protoc. Protein Sci. *91*, 19.23.1-19.23.15.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. Mol. Cell *25*, 15–30.

Sakaguchi, M., Cai, W., Wang, C.H., Cederquist, C.T., Damasio, M., Homan, E.P., Batista, T., Ramirez, A.K., Gupta, M.K., Steger, M., et al. (2019). FoxK1 and FoxK2 in insulin regulation of cellular and mitochondrial metabolism. Nat. Commun. *10*.

Sapcariu, S.C., Kanashova, T., Weindl, D., Ghelfi, J., Dittmar, G., and Hiller, K. (2014). Simultaneous extraction of proteins and metabolites from cells in culture. MethodsX *1*, 74–80.

Sanders, W.S., Bridges, S.M., McCarthy, F.M., Nanduri, B., Burgess, S.C. (2007) Prediction of peptides observable by mass spectrometry applied at the experimental set level, MBC Bioinformatics, 8, 1-9

Schulze, W.X., and Mann, M. (2004). A Novel Proteomic Screen for Peptide-Protein Interactions. J. Biol. Chem. *279*, 10756–10764.

Schulze, W.X., Deng, L., and Mann, M. (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol. Syst. Biol. *1*, E1–E13.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Seipel, K., Georgiev, O., and Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J. *11*, 4961–4968.

Selbach, M., Paul, F.E., Brandt, S., Guye, P., Daumke, O., Backert, S., Dehio, C., and Mann, M. (2009). Host Cell Interactome of Tyrosine-Phosphorylated Bacterial Proteins. Cell Host Microbe *5*, 397–403.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shen, N., Jiang, S., Lu, J., Yu, X., Lai, S., Zhang, J., Zhang, J., Tao, W., and Wang, X. (2015). The Constitutive Activation of Egr-1 / C / EBPa Mediates the Development of Type 2 Diabetes Mellitus by Enhancing Hepatic Gluconeogenesis. *185*.

Shiratori, T., Miyatake, S., Ohno, H., Nakaseko, C., Isono, K., Bonifacino, J.S., and Saito, T. (1997). Tyrosine phosphorylation controls internalization of CTLA-4 by regulating its interaction with clathrin-associated adaptor complex AP-2. Immunity *6*, 583–589.

Simone, A. De, Kitchen, C., Kwan, A.H., Sunde, M., Dobson, C.M., and Frenkel, D. (2012). Intrinsic disorder modulates protein self-assembly and aggregation. Proc. Natl. Acad. Sci. *109*, 6951–6956.

Slomiany, B.A., Arigo, K.L.D., Kelly, M.M., and Kurtz, D.T. (2000). C/EBPa Inhibits Cell Growth via Direct Repression of E2F-DP-Mediated Transcription. Mol. Cell. Biol. *20*, 5986–5997.

Smits, A.H., and Vermeulen, M. (2016). Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. Trends Biotechnol. *34*, 825–834.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. *5*, 699–711.

Stefano, B. Di, Sardina, J.L., Oevelen, C. Van, Collombet, S., Kallin, E.M., and Vicent, G.P. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

Stein, A., and Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. PLoS One *3*, 1–10.

Stoilova, B., Kowenz-Leutz, E., Scheller, M., and Leutz, A. (2013). Lymphoid to Myeloid Cell Trans-Differentiation Is Determined by C/EBPβ Structure and Post-Translational Modifications. PLoS One *8*, 1–9.

Subramanian, A., Subramanian, A., Tamayo, P., Tamayo, P., Mootha, V.K., Mootha, V.K., Mukherjee, S., Mukherjee, S., Ebert, B.L., Ebert, B.L., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Sukhatme, V.P., Cao, X.M., Chang, L.C., Tsai-Morris, C.H., Stamenkovich, D., Ferreira, P.C., Cohen, D.R., Edwards, S.A., Shows, T.B., and Curran, T. (1988). A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. Cell *53*, 37–43.

Sutherland, J.A., Cook, A., Bannister, A.J., and Kouzarides, T. (1992). Conserved motifs in Fos and Jun define a new class of activation domain. Genes Dev. *6*, 1810–1819.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Taatjes, D.J. (2017). Transcription Factor–Mediator Interfaces: Multiple and Multi-Valent. J. Mol. Biol. *429*, 2996–2998.

Thiel, G., and Cibelli, G. (2002). Regulation of life and death by the zinc finger transcription factor Egr-1. J. Cell. Physiol. *193*, 287–292.

Tinti, M., Kiemer, L., Costa, S., Miller, M., Sacco, F., Olsen, J. V., Carducci, M., Paoluzi, S., Langone, F., Workman, C.T., et al. (2014). The SH2 domain interaction landscape. Cell Rep. *3*, 1293–1305.

Tipton, J.D., Tran, J.C., Catherman, A.D., Ahlf, D.R., Durbin, K.R., and Kelleher, N.L. (2011). Analysis of intact protein isoforms by mass spectrometry. J. Biol. Chem. *286*, 25451–25458.

Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A Million peptide motifs for the molecular biologist. Mol. Cell *55*, 161–169.

Tontonoz, P., Hu, E., Graves, R.A., Budavari, A.I., and Spiegelman, B.M. (1994). mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev. *8*, 1224–1234.

Trinkle-Mulcahy, L. (2019). Recent advances in proximity-based labelling methods for interactome mapping [version 1; referees: 2 approved]. F1000Research *8*.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576.e16.

Tyers, M., and Mann, M. (2003). From genomics to proteomics. Nature *422*, 193–197.

Varadi, M., Zsolyomi, F., Guharoy, M., and Tompa, P. (2015). Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. PLoS One *10*, 1–16.

Varma, A.K., Brown, R.S., Birrane, G., and Ladias, J.A.A. (2005). Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex. Biochemistry *44*, 10941–10946.

Vojnic, E., Mourão, A., Seizl, M., Simon, B., Wenzeck, L., Larivière, L., Baumli, S., Baumgart, K., Meisterernst, M., Sattler, M., et al. (2011). Structure and VP16 binding of the Mediator Med25 activator interaction domain. Nat. Struct. Mol. Biol. *18*, 404–410.

Volkmer, R., Tapia, V., and Landgraf, C. (2012). Synthetic peptide arrays for investigating protein interaction domains. FEBS Lett. *586*, 2780–2786.

Wang, H., Iakova, P., Wilde, M., Welm, A., Goode, T., Roesler, W.J., and Timchenko, N.A. (2001). C/EBPa Arrests Cell Proliferation through Direct Inhibition of Cdk2 and Cdk4. Mol. Cell *8*, 817–828.

Wang, N., Finegold, M.J., Bradley, A., Ou, C.N., Abdelsayed, S. V, Wilde, M.D., Taylor, L.R., Wilson, D.R., and Darlington, G.J. (1995). Impaired Energy Homeostasis in C / EBPa Knockout Mice. Science (80). *269*, 1108–1112.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J. Mol. Biol. *337*, 635–645.

Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc. Natl. Acad. Sci. U. S. A. *105*, 5762–5767.

Woodsmith, J., Casado-Medrano, V., Benlasfer, N., Eccles, R.L., Hutten, S., Heine, C.L., Thormann, V., Abou-Ajram, C., Rocks, O., Dormann, D., et al. (2018). Interaction modulation through arrays of clustered methyl-arginine protein modifications. Life Sci. Alliance *1*, e201800178.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol. *293*, 321–331.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. *16*, 18–29.

Wu, M., Li, C., and Zhu, X. (2018). FLT3 inhibitors in acute myeloid leukemia. J. Hematol. Oncol. *11*, 1–10.

Wu, Z., Cheng, Z., Sun, M., Wan, X., Liu, P., He, T., Tan, M., and Zhao, Y. (2015). A Chemical Proteomics Approach for Global Analysis of Lysine Monomethylome Profiling. Mol. Cell. Proteomics *14*, 329–339.

Xu, P., Balczerski, B., Ciozda, A., Louie, K., Oralova, V., Huysseune, A., and Gage Crump, J. (2018). Fox proteins are modular competency factors for facial cartilage and tooth specification. Dev. *145*.

Yates J.R. (1998), Mass Spectrometry and the Age of the Proteome, Journal of Mass Spectrometry, *33*, 1-19

Yang, Z., Drumea, K., Cormier, J., Wang, J., Zhu, X., and Rosmarin, A.G. (2011). GABP transcription factor is required for myeloid differentiation , in part , through its control of Gfi-1 expression. Blood *118*, 2243–2254.

Yu, J., Baron, V., Mercola, D., Mustelin, T., and Adamson, E.D. (2007). A network of p73 , p53 and Egr1 is required for efficient apoptosis in tumor cells. 436–446.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors : establishing competence for gene expression Parameters affecting transcription factor access to target sites in chromatin Initiating events in chromatin : pioneer factors bind first. Genes Dev. 2227–2241.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signalling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc. Natl. Acad. Sci. U. S. A. *94*, 569–574.

Zhang, F., Lin, M., Abidi, P., Thiel, G., and Liu, J. (2003). Specific Interaction of Egr1 and c/EBPβ Leads to the Transcriptional Activation of the Human Low Density Lipoprotein Receptor Gene. J. Biol. Chem. *278*, 44246–44254.

Zhang, O., Kay, L.E., Olivier, J.P., and Forman-Kay, J.D. (1994). Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J. Biomol. NMR *4*, 845–858.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., et al. (2004). Enhancement of haematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα. Immunity *21*, 853–863.

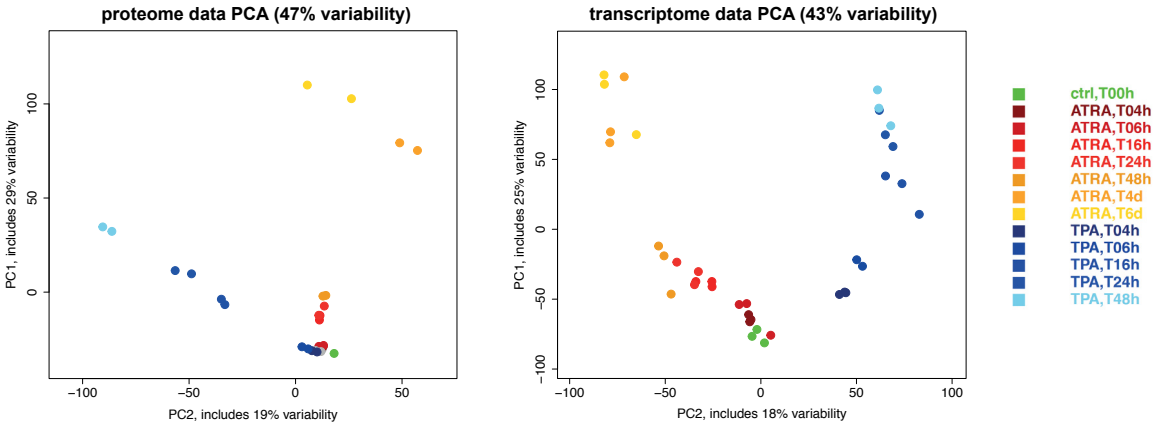
Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., and Yates, J.R. (2013). Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. *113*, 2343–2394.

Zheng, X., Ravatn, R., Lin, Y., Shih, W.-C., Rabson, A., Strair, R., Huberman, E., Conney, A., and Chin, K.-V. (2002). Gene expression of TPA induced differentiation in HL-60 cells by DNA microarray analysis. Nucleic Acids Res. *30*, 4489–4499.

Zhou, J., Zhao, S., and Dunker, A.K. (2018). Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signalling and Regulation. J. Mol. Biol. *430*, 2342–2359.

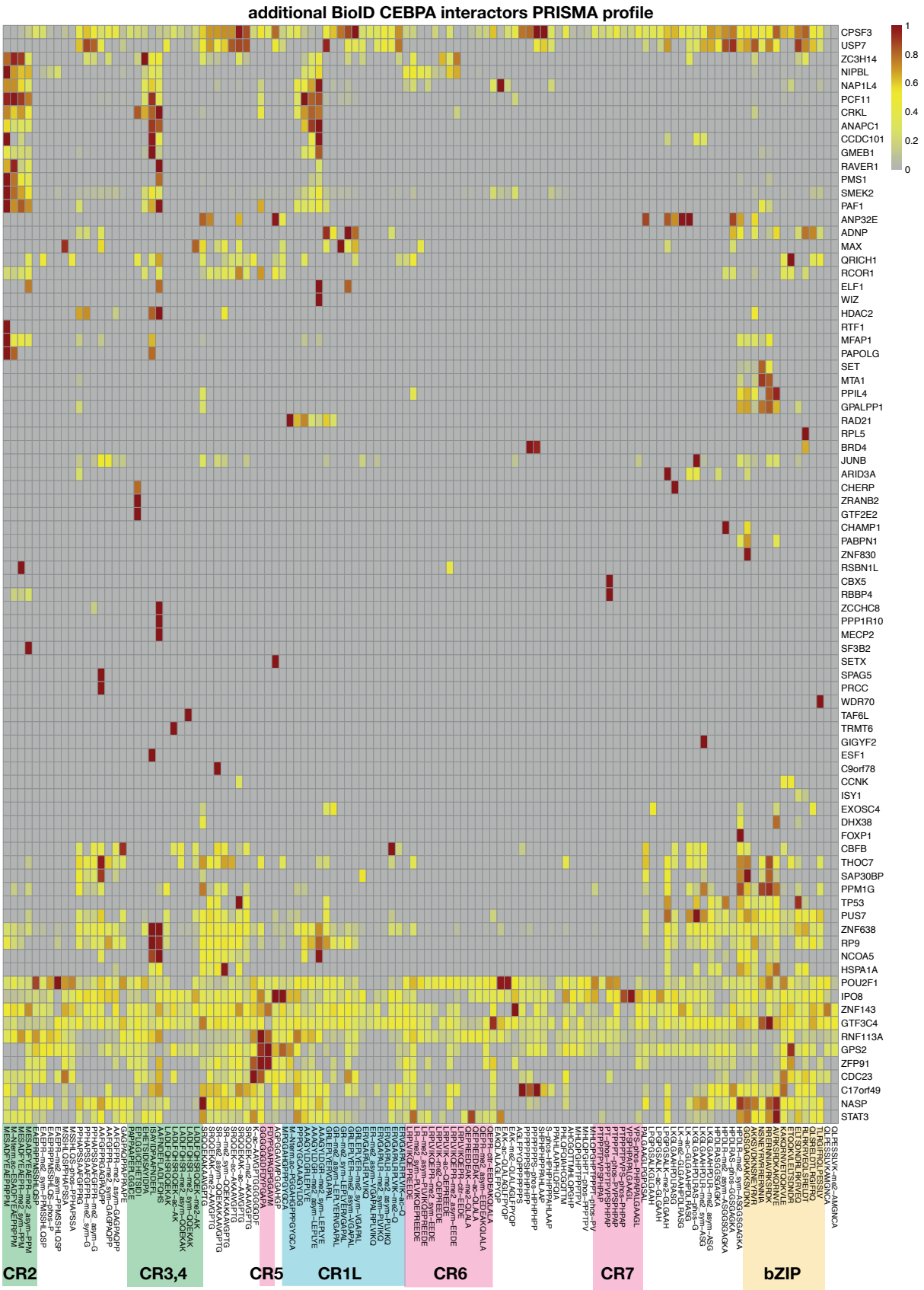
Zubarev, R.A., and Makarov, A. (2013). Orbitrap mass spectrometry. Anal. Chem. *85*, 5288–5296.

Supplementary Figures and Tables

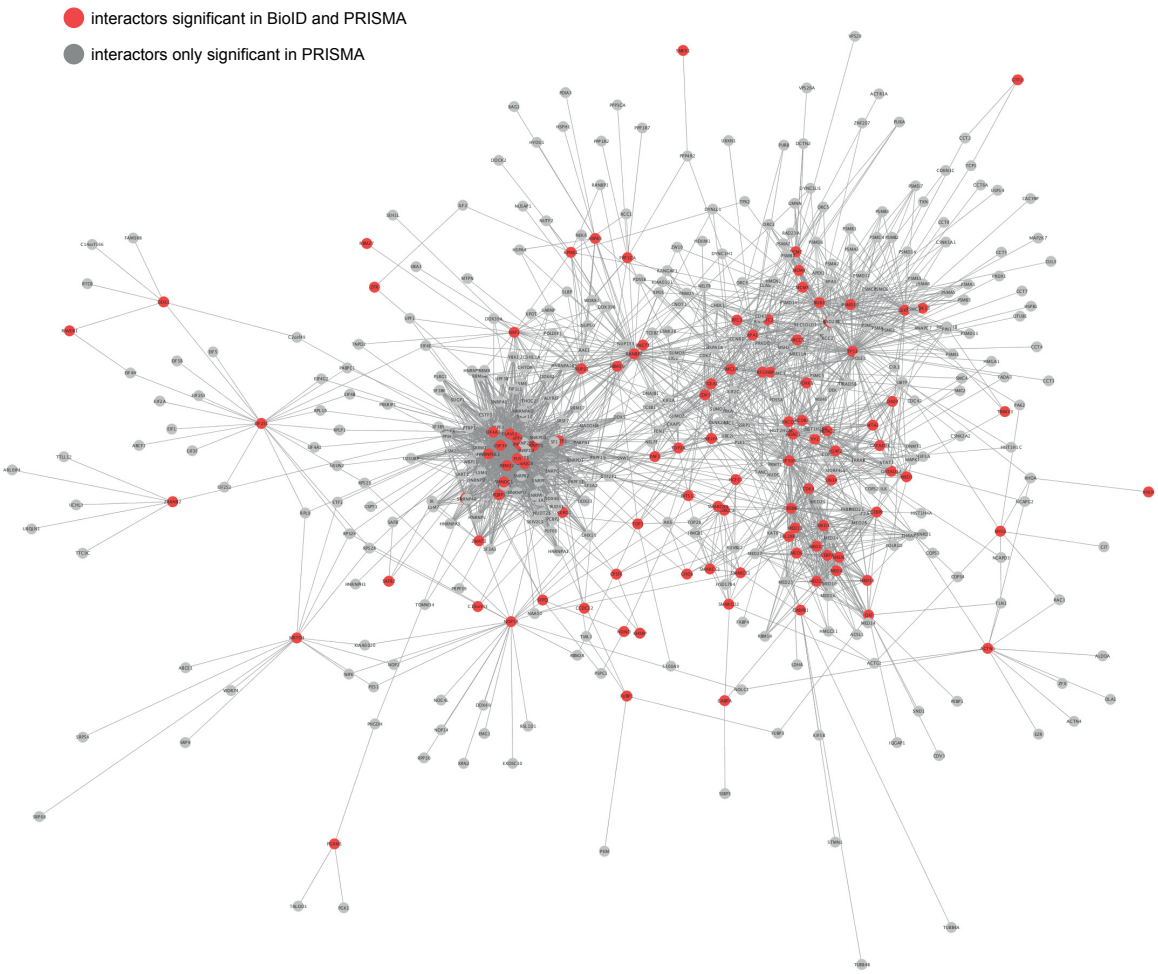


Supplemental Figure 1: Principal Component Analysis of NB4 differentiation. NB4 cells were treated with ATRA or TPA for the indicated time points. Transcriptome and proteome were analysed with microarray and SILAC based proteomic respectively.

Supplemental Figure 2: PRISMA binding profile of validated C/EBPα interactors. Heatmap displays validated C/EBPα interactors (significant in PRISMA and BioID or databases). Normalised LFQ intensities of proteins (x axis) are displayed across C/EBPα PRISMA peptides ordered from N- to C- terminus.



Supplemental Figure 3: PRISMA binding profile of C/EBPα BioID interactors not significant in PRISMA. Heatmap displays C/EBPα interactors significant in BioID and with an FDR in PRISMA < 10%. Normalised LFQ intensities of proteins (x axis) are displayed across C/EBPα PRISMA peptides ordered from N- to C- terminus.



Supplemental Figure 4: C/EBPα interactors significant only in PRISMA are connected to validated C/EBPα interactors in a STRING network.
Protein interaction network visualised with the Cytoscape software. C/EBPα interactors significant in both BioID and PRISMA are depicted as red nodes. C/EBPα interactors significant only in PRISMA are depicted as grey nodes. Experimentally validated interactions with a confidence score > 0.6 were retrieved from the STRING database and are depicted as edges.

Supplemental Figure 5: C/EBPα interactors significant only in BioID are connected to validated C/EBPα interactors in a STRING network.
Protein interaction network visualised with the Cytoscape software. C/EBPα interactors significant in both BioID and PRISMA are depicted as red nodes. C/EBPα interactors significant only in BioID are depicted as grey nodes. Experimentally validated interactions with a confidence score > 0.6 were retrieved from the STRING database and are depicted as edges.

Gene.name	uniprot	FDR	ratio.cebpa.bioid	rank
CEBPA	P49715	4.19E-05	10.32	1
HOXD13	P35453	2.36E-05	6.27	2
CEBPE	Q15744	4.78E-05	6.46	3
CBFB	Q13951-2	8.00E-06	4.68	4
CEBPD	P49716	2.34E-04	6.28	5
NFIL3	Q16649	3.05E-05	4.97	6
NCOA1	Q15788-3	2.34E-04	6.11	7
TCF20	Q9UGU0-2	2.34E-04	6.01	8
TET2	Q6N021	2.34E-05	4.67	9
ARID1B	Q8NFD5	2.02E-04	5.60	10
HOXA13	P31271	1.44E-06	3.51	11
BCAR2	Q05GC8	1.80E-04	5.42	12
FLI1	Q01543	3.05E-05	4.43	13
CEBPB	P17676	2.79E-04	5.57	14
NCOA3	Q9Y6Q9-5	3.40E-04	5.68	15
CREBBP	Q92793	3.05E-05	4.26	16
SIRT1	Q96EB6	4.19E-05	4.37	17
RUNX2	Q13950-2	5.75E-05	4.45	18
NCOA2	Q15596	5.70E-04	5.80	19
ATF3	P18847-5	4.19E-05	4.29	20
DACH1	Q9UI36-2	5.17E-05	4.38	21
ZNF292	J3KNV1	2.19E-04	5.04	22
CEBPG	P53567	3.05E-05	3.98	23
IPO7	O95373	5.19E-05	4.19	24
CHD7	Q9P2D1	3.85E-04	5.25	25
FOXP1	Q9H334	1.30E-04	4.49	26
KMT2C	Q8NEZ4	4.39E-05	3.85	27
CBFB	Q13951	5.52E-05	3.91	28
JDP2	Q8WYK2	2.85E-06	2.93	29
SATB2	C9JR56	3.47E-05	3.64	30
QSER1	Q2KHR3	4.39E-05	3.70	31
IKZF1	Q13422	3.05E-05	3.57	32
ARID3B	Q8IVW6-4	4.68E-04	4.80	33
BCOR	Q6W2J9	5.00E-04	4.80	34
MAML1	Q92585	1.80E-04	4.21	35
TLE3	Q04726	7.38E-04	4.98	36
EP300	Q09472	7.62E-04	4.98	37
JUN	P05412	5.18E-05	3.57	38
HCFC1	P51610	7.33E-05	3.66	39
NCOR2	Q9Y618-5	9.23E-04	4.97	40
ZNF281	Q9Y2X9	5.66E-05	3.50	41
ZEB2	O60315-2	1.67E-03	5.32	42
CUX1	P39880-3	2.71E-04	4.12	43
BCL11A	Q9H165	3.81E-04	4.26	44
GSE1	Q14687	5.09E-04	4.32	45
FOSL2	P15408	6.54E-04	4.39	46
PPARG	P37231-2	2.83E-05	3.05	47
PRR12	Q9ULL5-3	3.72E-04	4.01	48
MYB-myb	P10242-10	2.19E-04	3.71	49
ZBTB34	Q8NCN2	5.18E-05	3.16	50
TFAP2A	P05549-5	2.36E-05	2.91	51
ARID3A	Q99856	4.68E-04	4.03	52
ZNF516	Q92618	1.67E-04	3.47	53
ARNT	P27540-2	1.45E-04	3.37	54
ATF2	B8ZZU6	2.55E-04	3.60	55
ZNF384	Q8TF68	1.81E-04	3.45	56
ZBTB3	Q9H5J0	7.31E-05	3.08	57
RUNX1	Q01196-8	1.80E-04	3.38	58
RAI1	Q7Z5J4	1.25E-03	4.30	59
AHDC1	Q5TGY3	9.75E-04	4.14	60
KMT2D	O14686	3.31E-04	3.58	61
UBN2	Q6ZU65	8.92E-04	4.07	62
RREB1	Q92766-2	3.34E-04	3.57	63
TRIM33	Q9UPN9	7.82E-04	3.98	64
FOSL1	P15407	1.62E-04	3.25	65
EGR1	P18146	1.95E-04	3.30	66
CREB1	P16220	2.59E-04	3.41	67
FOXJ3	Q9UPW0-2	1.62E-04	3.20	68
RLF	Q13129	2.34E-04	3.32	69
JUNB	P17275	2.74E-03	4.69	70
IRF2BP2	Q7Z5L9	6.97E-05	2.87	71
BAZ2B	Q9UIF8-4	1.53E-05	2.45	72
NR2C2	P49116	1.67E-04	3.11	73
NCOR1	O75376	9.92E-04	3.87	74
DPF2	Q92785	4.72E-05	2.66	75
KIAA1551	Q9HCM1	1.67E-04	3.03	76
ARID1A	O14497	2.71E-04	3.20	77
VEZF1	J3QSH4	3.40E-04	3.28	78
TFAP4	Q01664	4.37E-04	3.36	79
ELMSAN1	A0A1C7CYX1	2.27E-03	4.25	80
LCORL	Q8N3X6	1.78E-03	3.99	81
MEF2A	Q02078-3	4.54E-04	3.26	82
NIPBL	Q6KC79-2	2.69E-05	2.39	83
DNTTIP1	Q9H147	1.55E-03	3.85	84
NFRKB	Q6P4R8-3	2.15E-03	4.05	85
MLLT6	A0A087WW39	1.62E-04	2.81	86
ZNF609	O15014	2.04E-03	3.95	87
PIAS1	O75925	4.68E-04	3.19	88
PLAGL2	Q9UPG8	1.67E-04	2.82	89
KDM6A	O15550	2.71E-04	2.96	90
TCF3	P15923	2.69E-04	2.93	91
PHF21A	Q96BD5-2	8.31E-04	3.38	92
ZNF362	Q5T0B9	1.12E-03	3.53	93
NACC1	Q96RE7	1.35E-04	2.68	94

Gene.name	uniprot	FDR	ratio.cebpa.bioid	rank
IPO8	O15397	1.62E-04	2.74	95
LIN54	Q6MZP7	2.37E-03	3.94	96
GABPA	Q06546	6.52E-04	3.17	97
ZNF217	O75362	2.36E-03	3.81	98
TGIF1	Q15583-4	4.02E-05	2.27	99
TRIM24	O15164	2.45E-03	3.82	100
SMARCE1	Q969G3-2	4.39E-05	2.28	101
IKZF5	Q9H5V7	7.58E-05	2.41	102
MED15	Q96RN5-3	9.15E-04	3.23	103
TBL1XR1	Q9BZK7	2.27E-03	3.63	104
NCOA6	Q14686	2.23E-03	3.58	105
SPEN	Q96T58	2.58E-03	3.66	106
ETV6	P41212	2.04E-03	3.52	107
HIVEP1	F5H212	9.49E-04	3.13	108
IRF2BP1	Q8IU81	3.02E-04	2.67	109
ATF7	P17544-4	1.56E-03	3.34	110
GLTSCR1	Q9NZM4-2	1.45E-03	3.30	111
SMARCC1	Q92922	4.39E-05	2.14	112
PITX1	P78337	1.70E-04	2.46	113
BACH1	O14867	2.34E-04	2.49	114
HOMEZ	Q8IX15	5.00E-04	2.74	115
ZMYM4	Q5VZL5	1.13E-04	2.27	116
ETS1	P14921	6.62E-04	2.79	117
FOXK2	Q01167	3.95E-04	2.61	118
EBF3BF1	Q9H4W6	8.33E-04	2.82	119
SAP130	H7BXF5	4.61E-04	2.60	120
ZNF845	Q96IR2	4.16E-04	2.52	121
EP400	Q96L91-3	8.59E-04	2.76	122
RARA	P10276-2	1.13E-03	2.85	123
ZNF148	Q9UQR1	1.13E-03	2.78	124
LIN37	Q96GY3	1.13E-03	2.76	125
NR2C1	P13056	4.18E-03	3.41	126
SMARCC2	Q8TAQ2-2	5.17E-05	1.88	127
OGT	O15294	1.70E-03	2.91	128
MYBL2	P10244	5.60E-03	3.58	129
MAFK	O60675	5.00E-04	2.41	130
MAX	P61244-2	3.53E-03	3.19	131
RBAK	Q9NYW8	1.18E-03	2.65	132
MAFG	O15525	1.59E-03	2.76	133
SMARCA2	P51531-2	5.18E-05	1.79	134
WT1	H0Y7K5	2.69E-04	2.14	135
RFX1	P22670	2.25E-03	2.87	136
JUND	P17535	4.61E-03	3.18	137
BCL7B	Q9BQE9-4	1.18E-04	1.84	138
BRD4	O60885	5.43E-04	2.20	139
PPIL2	Q13356	4.44E-04	2.14	140
ATF1	P18846	5.36E-03	3.14	141
JMJD1C	Q15652	1.67E-03	2.55	142
C18orf25RK	Q96B23-2	3.73E-04	2.06	143
GOPC	Q9HD26-2	3.24E-03	2.83	144
RAD54L2	Q9Y4B4	1.02E-02	3.53	145
GFI1	Q99684	1.59E-03	2.51	146
GABPB1	Q06547-2	5.19E-04	2.13	147
ELF1KFZp68	P32519	6.52E-04	2.19	148
PRRC2B	Q5JSZ5	3.31E-03	2.80	149
SMARCA4	P51532-5	6.33E-05	1.63	150
ZNF655	Q8N720	2.80E-03	2.66	151
ZNF316	A6NFI3	1.59E-03	2.41	152
FOXK1	P85037	2.69E-04	1.88	153
DGCR14	Q96DF8	2.04E-03	2.47	154
ZNF121	P58317	2.04E-03	2.47	155
POU2F1	P14859-5	2.66E-03	2.56	156
USF2	Q15853-2	8.83E-04	2.15	157
HDAC4	P56524	5.97E-03	2.95	158
GLTSCR1L	Q6AI39	2.43E-03	2.50	159
CHD8	Q9HCK8	1.12E-03	2.21	160
ZKSCAN4	Q969J2	1.58E-03	2.32	161
GPS2	I3L4X7	3.84E-03	2.65	162
C9orf78	Q9NZ63	2.20E-04	1.74	163
MED1	Q15648	3.06E-04	1.81	164
MEF2D	Q14814-4	2.14E-03	2.37	165
ZSCAN21	Q9Y5A6	3.84E-03	2.61	166
SMARCD2	J3KMX2	5.34E-05	1.46	167
L3MBTL3	Q96JM7-2	4.64E-03	2.65	168
RFX5	P48382	2.47E-03	2.35	169
MED6	O75586	2.71E-04	1.71	170
RPA1	P27694	2.58E-03	2.27	171
POGZ	Q7Z3K3-5	7.31E-03	2.75	172
BEND3	Q5T5X7	1.54E-03	2.08	173
ANP32E	Q9BTT0	2.34E-04	1.60	174
ANLN	Q9NQW6	5.25E-04	1.77	175
E2F8	A0AVK6	1.78E-03	2.11	176
KMT2B	Q9UMN6	1.78E-03	2.10	177
PAXIP1	Q6ZW49	2.17E-03	2.13	178
IKZF1	Q13422-7	1.17E-02	2.89	179
SMARCD1	Q96GM5	9.56E-04	1.85	180
ZNF24	P17028	1.88E-03	2.04	181
ZNF143	P52747-2	5.00E-03	2.41	182
ZMIZ2	Q8NF64-3	4.55E-03	2.36	183
ZMYM2	Q9UBW7	3.48E-04	1.60	184
WIZ	O95785-3	1.08E-02	2.63	185
ADNP	Q9H2P0	5.29E-04	1.56	186
NAB2	Q15742	7.44E-03	2.39	187
HDAC3	O15379	5.64E-03	2.23	188

Gene.name	uniprot	FDR	ratio.cebpa.bioid	rank
LIN9	Q5TKA1	2.06E-02	2.96	189
CDK8	P49336-2	7.75E-03	2.35	190
GATAD2B	Q8WXI9	5.70E-04	1.53	191
MED22	E9PGW7	4.52E-04	1.48	192
KIAA0907	Q7Z7F0	4.68E-04	1.48	193
C20orf112	Q96MY1	3.20E-03	1.95	194
SMARCB1	Q12824	2.31E-03	1.84	195
INTS12	Q96CB8	5.00E-04	1.47	196
ZNF318	Q5VUA4	2.16E-03	1.81	197
MNT	Q99583	1.32E-02	2.55	198
BCORL1	Q5H9F3-3	3.37E-02	3.25	199
ZBTB21	Q9ULJ3	2.19E-02	2.87	200
LRIF1	Q5T3J3	4.64E-03	2.04	201
FLYWCH2	Q96CP2	2.00E-04	1.29	202
SETX	Q7Z333	5.30E-03	2.08	203
GMEB2	Q9UKD1	1.43E-02	2.54	204
LENG8	Q96PV6-2	9.81E-03	2.31	205
RNF113A	O15541	1.67E-03	1.67	206
NAP1L4	Q99733	2.15E-03	1.73	207
ZFP91FP91-CNTF	Q96JP5-2	7.90E-03	2.19	208
PPIL1	Q9Y3C6	9.62E-04	1.52	209
ISL2	Q96A47	1.25E-02	2.36	210
MED17	Q9NVC6	5.40E-03	1.97	211
KDM3B	Q7LBC6	3.26E-03	1.79	212
ZBTB9	Q96C00	8.76E-03	2.15	213
ASXL2	Q76L83	1.83E-03	1.60	214
RPL5	P46777	1.70E-04	1.15	215
MED20	Q9H944	4.19E-03	1.82	216
CHD3	Q12873	1.25E-02	2.25	217
MTA3	E7EQY4	1.14E-03	1.46	218
KMT2A	Q03164-2	1.63E-02	2.39	219
STAG2	Q8N3U4	2.03E-02	2.52	220
ZMYND8	Q9ULU4-12	7.08E-03	1.98	221
ZC3H10	Q96K80	2.94E-02	2.78	222
ZNF687	Q8N1G0	6.11E-03	1.92	223
TNRC18	O15417-2	1.08E-02	2.15	224
ELF2	Q15723-2	5.55E-04	1.28	225
EYA3	Q99504-5	1.80E-02	2.37	226
BRD9	Q9H8M2	1.42E-02	2.24	227
ZNF131	P52739-2	2.09E-02	2.45	228
C15orf39	Q6ZRI6-2	2.09E-02	2.44	229
PBX2	P40425	1.41E-02	2.22	230
PHF12	Q96QT6	1.03E-02	2.07	231
PPIL3	Q9H2H8	9.58E-03	2.00	232
PRDM15NF298	E7ER26	4.07E-03	1.65	233
EHMT1	Q9H9B1	1.28E-02	2.06	234
ZNF184	Q99676	1.97E-02	2.28	235
MAML2	Q8IZL2	2.55E-03	1.50	236
ZBTB40	Q9NUA8	1.83E-02	2.23	237
CHD9	Q3L8U1-2	1.27E-02	2.04	238
ZFAND3	Q9H8U3	5.86E-03	1.70	239
EMSY11orf30	Q7Z589	4.67E-03	1.62	240
MBD2	Q9UBB5	7.06E-04	1.18	241
ZNF174	Q15697	2.31E-02	2.27	242
CIC	I3L2J0	3.44E-02	2.53	243
FBRSL1	Q9HCM7	2.47E-02	2.29	244
ZHX3	Q9H4I2	3.56E-02	2.53	245
ACTL6A	O96019	1.22E-03	1.25	246
THOC7	Q6I9Y2	3.08E-02	2.40	247
CCNK	O75909	8.64E-03	1.71	248
RCOR1	J3KN32	9.28E-04	1.15	249
ZNF644	Q9H582	2.92E-02	2.27	250
PML	P29590	4.86E-03	1.49	251
TAF4	O00268	2.92E-02	2.25	252
GATAD2A	Q86YP4	1.94E-03	1.27	253
HCFC1	P51610-4	3.11E-03	1.36	254
SLU7	O95391	1.56E-02	1.89	255
EPC2	Q52LR7	9.28E-03	1.68	256
BCL9	O00512	1.18E-02	1.75	257
TLK2	Q86UE8-2	4.61E-04	1.01	258
CHD6	Q8TD26	2.46E-02	2.09	259
RPUSD2	Q8IZ73	5.60E-03	1.48	260
TCEA1	P23193	6.19E-04	1.03	261
YEATS2	Q9ULM3	3.04E-02	2.18	262
PTPN22	Q9Y2R2-6	1.32E-02	1.76	263
CASP1	P29466-2	3.67E-03	1.35	264
IRF2BP2	Q7Z5L9-2	3.95E-02	2.33	265
CDC23	Q9UJX2	6.30E-03	1.48	266
WAC	Q9BTA9	2.86E-02	2.10	267
PHF2	O75151	2.72E-03	1.26	268
CCNT1	O60563	1.25E-02	1.68	269
ZNF3	P17036	1.35E-02	1.70	270
KDM3A	Q9Y4C1	3.82E-02	2.23	271
NELFA	Q9H3P2	1.06E-02	1.59	272
NAP1L1	P55209-2	2.77E-02	1.99	273
TAF6	P49848	4.71E-03	1.33	274
TDP2	O95551	7.39E-03	1.45	275
CDK9	P50750	1.27E-03	1.06	276
KANSL3	Q9P2N6-3	2.77E-02	1.94	277
MED12	Q93074-3	2.37E-02	1.84	278
QRICH1	Q2TAL8	1.32E-02	1.59	279
DEAF1	O75398	2.86E-02	1.90	280
SP4	Q02446	7.06E-03	1.35	281
C10orf12	A0A1B0GUH9	3.05E-02	1.92	282
ZNF652	Q9Y2D9	1.28E-02	1.53	283
ZBTB39	O15060	2.54E-02	1.82	284
CASC5	Q8NG31-2	1.61E-02	1.61	285
YY1	P25490	3.01E-02	1.87	286
IKZF3	Q9UKT9	3.54E-02	1.96	287
MSANTD2	Q6P1R3-3	8.96E-03	1.38	288
PCF11	O94913	3.73E-02	1.96	289
RBBP4	Q09028-3	2.59E-03	1.07	290
SET	Q01105	4.52E-03	1.16	291
SMEK2	Q5MIZ7-3	1.40E-02	1.46	292
ZMAT2	Q96NC0	1.83E-02	1.55	293
ZBTB1	Q9Y2K1-2	1.43E-02	1.45	294
PSMC3IP	Q9P2W1	1.00E-02	1.34	295
GTF3C1	Q12789	4.03E-02	1.92	296
CCDC101	Q96ES7	2.69E-02	1.69	297
SF3B4	Q15427	4.84E-02	2.02	298
POLK	Q9UBT6-3	1.35E-02	1.40	299
TCERG1	O14776-2	1.11E-02	1.33	300
FBRS	J3KNZ9	3.28E-02	1.74	301
FOXC1	Q12948	7.39E-03	1.21	302
TAF10	Q12962	6.45E-03	1.17	303
UNCX	A6NJT0	4.42E-03	1.09	304
DCUN1D5	Q9BTE7	1.18E-02	1.33	305
TLK1	Q9UKI8-5	1.45E-02	1.38	306
CCDC174	Q6PII3	3.82E-02	1.79	307
CDK13	Q14004	1.03E-02	1.27	308
PIAS2	O75928-2	2.92E-02	1.63	309
GMEB1	Q9Y692-2	2.57E-02	1.57	310
CCNT2	O60583-2	1.06E-02	1.26	311
MED4	Q9NPJ6	9.58E-03	1.23	312
BPTF	Q12830-4	1.70E-02	1.40	313
CDCA5	Q96FF9	1.35E-02	1.33	314
AGGF1	Q8N302	1.48E-02	1.34	315
PQBP1	O60828-2	3.76E-02	1.71	316
DPF1	C8C3P2	1.65E-02	1.36	317
MTA1	Q13330-3	7.32E-03	1.13	318
MED8	Q96G25	1.19E-02	1.24	319
RBM5	P52756	4.22E-02	1.71	320
SKP1	P63208	2.80E-02	1.48	321
RBM33	Q96EV2	2.19E-02	1.37	322
HMG20B	Q9P0W2	3.28E-02	1.53	323
USF1	B1AQP1	4.06E-02	1.60	324
KIAA1704	A0A0A0MRI1	1.34E-02	1.18	325
RAVER1	A0A087WZ13	9.75E-03	1.08	326
BRCA1	P38398	3.98E-02	1.55	327
ZEB1	P37275-3	1.10E-02	1.11	328
RBM27	Q9P2N5	1.08E-02	1.09	329
C19orf43	Q9BQ61	1.70E-02	1.19	330
KNSTRN	V9GY01	4.50E-02	1.54	331
PPWD1	Q96BP3	1.12E-02	1.05	332
MBD3	O95983	1.17E-02	1.06	333
ZNF414	Q96IQ9-2	1.35E-02	1.08	334
SATB1	Q01826	2.77E-02	1.27	335
SS18L2	Q9UHA2	1.97E-02	1.14	336
KLF3	P57682	2.31E-02	1.18	337
VGLL4	Q14135-4	3.59E-02	1.32	338
MEF2A	Q02078-4	4.80E-02	1.44	339
RMI1	Q9H9A7	2.79E-02	1.19	340
DIDO1	Q9BTC0	1.70E-02	1.03	341
IFT57	Q9NWB7	4.06E-02	1.31	342
AASDH	Q4L235-3	4.51E-02	1.31	343
HDAC2	Q92769	2.53E-02	1.09	344
RBM10	A0A0A0MR66	2.56E-02	1.04	345
LCOR	Q96JN0-2	4.84E-02	1.24	346
ZNF141	Q15928	3.13E-02	1.06	347
THAP4	Q8WY91	4.35E-02	1.17	348
C1orf52	Q8N6N3	4.11E-02	1.13	349
UVSSA	Q2YD98	3.83E-02	1.07	350
ZRANB2	O95218-2	4.00E-02	1.08	351
CHD4	F5GWX5	3.50E-02	1.02	352
TAF9	Q16594	4.70E-02	1.08	353
PAF1	Q8N7H5-3	4.34E-02	1.01	354

Supplemental Table 1: Ranked list of P42-C/EBPα interactors detected by BioID experiments in NB4 cells. Proteins are ranked by enrichment and FDR.

GENE	UNIPROT	CEBPA isoform specificity	ratio.p42.p30	FDR.p42.p30
GLTSCR1	Q9NZM4	p42	0.79	6.57E-02
MED6	O75586	p42	1.06	2.45E-02
ZBTB34	Q8NCN2	p42	0.99	5.58E-02
E2F8	A0AVK6	p42	0.94	2.52E-02
SATB2	C9JR56	p42	2.11	7.00E-02
PPARG	P37231	p42	3.78	1.78E-02
SMARCC2	Q8TAQ2	p42	0.82	7.19E-02
ARID1A	O14497	p42	1.14	4.33E-02
TRIM24	O15164	p42	0.90	3.55E-02
OGT	O15294	p42	1.06	6.57E-02
ZEB2	O60315	p42	0.82	9.76E-02
CCNT1	O60563	p42	0.55	7.00E-02
ZNF217	O75362	p42	1.08	4.72E-02
ACTL6A	O96019	p42	0.72	9.76E-02
ANXA1	P04083	p42	1.72	2.39E-02
JUN	P05412	p42	1.09	4.81E-02
TFAP2A	P05549	p42	2.66	1.91E-02
ETS1	P14921	p42	1.10	7.74E-02
FOSL1	P15407	p42	0.90	4.81E-02
FOSL2	P15408	p42	1.29	2.45E-02
JUNB	P17275	p42	1.08	3.43E-02
EGR1	P18146	p42	5.10	7.22E-04
ATF3	P18847	p42	1.49	1.57E-02
ARNT	P27540	p42	1.21	7.00E-02
RPA1	P27694	p42	1.19	8.55E-02
CASP1	P29466	p42	1.38	7.00E-02
HOXA13	P31271	p42	1.29	8.74E-02
HOXD13	P35453	p42	1.38	5.35E-02
CDK8	P49336	p42	1.15	5.12E-02
SMARCA4	P51532	p42	0.79	4.38E-02
HDAC4	P56524	p42	0.86	7.74E-02
PITX1	P78337	p42	1.01	2.45E-02
FOXK1	P85037	p42	0.90	5.58E-02
FLI1	Q01543	p42	0.70	4.33E-02
EP300	Q09472	p42	1.57	2.39E-02
FOXC1	Q12948	p42	2.38	2.39E-02
RUNX2	Q13950	p42	1.43	4.12E-02
CBFB	Q13951	p42	1.02	9.73E-02
NCOA2	Q15596	p42	1.35	2.63E-02
MED1	Q15648	p42	0.75	3.52E-02
NAB2	Q15742	p42	0.95	6.48E-02
NCOA1	Q15788	p42	1.87	1.78E-02
NFIL3	Q16649	p42	2.24	1.57E-02
QSER1	Q2KHR3	p42	0.84	8.77E-02
AASDH	Q4L235	p42	1.06	1.78E-02
PRRC2B	Q5JSZ5	p42	1.45	4.15E-02
TET2	Q6N021	p42	1.68	1.78E-02
IRF2BP2	Q7Z5L9	p42	0.43	7.12E-02
IRF2BP1	Q8IU81	p42	0.97	4.38E-02
ARID3B	Q8IVW6	p42	0.74	3.43E-02
ARID1B	Q8NFD5	p42	0.95	4.81E-02
MAML1	Q92585	p42	1.10	1.57E-02
ZNF516	Q92618	p42	1.27	1.78E-02
DPF2	Q92785	p42	1.14	5.88E-02
CREBBP	Q92793	p42	2.01	1.78E-02
SMARCC1	Q92922	p42	0.82	4.38E-02
MED12	Q93074	p42	1.34	2.39E-02
SMARCE1	Q969G3	p42	0.90	4.21E-02
MED15	Q96RN5	p42	1.55	3.84E-02
ARID3A	Q99856	p42	0.79	2.61E-02
MED20	Q9H944	p42	1.00	1.87E-02
KIAA1551	Q9HCM1	p42	2.89	5.87E-02
ANLN	Q9NQW6	p42	0.88	2.78E-02
MED17	Q9NVC6	p42	0.98	6.57E-02
RBM22	Q9NW64	p42	0.50	3.55E-02
IFT57	Q9NWB7	p42	1.90	5.12E-02
CHD7	Q9P2D1	p42	1.06	4.27E-02
TRIM33	Q9UPN9	p42	1.61	6.83E-03
ZNF281	Q9Y2X9	p42	0.73	7.00E-02
NCOA3	Q9Y6Q9	p42	1.41	4.27E-02

GENE	UNIPROT	CEBPA isoform specificity	ratio.p42.p30	FDR.p42.p30
EHMT2	Q96KQ7	p30	-0.61	3.28E-02
GATA1	P15976	p30	-0.83	7.00E-02
BLM	P54132	p30	-0.61	7.12E-02
TFAP4	Q01664	p30	-0.68	2.61E-02
CENPC1	Q03188	p30	-1.01	4.33E-02
ZNF362	Q5T0B9	p30	-1.11	8.55E-02
AHDC1	Q5TGY3	p30	-1.16	6.93E-02
BCL11A	Q9H165	p30	-1.04	4.27E-02
EBF3BF1	Q9H4W6	p30	-0.97	8.49E-02

Supplemental Table 2: C/EBPα isoform-specific interactors detected by BioID experiments in NB4 cells.

Talks and Posters

Evelyn Ramberger, Daniel Perez-Hernandez, Valeriia Sapozhnikova, Elisabeth Kowenz-Leutz, Philipp Mertins, Gunnar Dittmar & Achim Leutz, **Peptide array protein pulldowns and BioID unravel SLiM based protein interactions in the transcription factor C/EBPα.** Abstract talk, European Proteomics Association (EuPa) Conference, March 24-28 2019, Potsdam, Germany

Evelyn Ramberger, Daniel Perez-Hernandez, Petr-Nazarov, Valeriia Sapozhnikova, Philipp Mertins, Gunnar Dittmar & Achim Leutz. **The C/EBPα interactome unravelled by a novel mass spectrometric screening method.** Poster, European Proteomics Association (EuPa) Conference, June 16-20 2018, Santiago de Compostela, Spain. **Poster award**

Publications

Lupberger J.*, Croonenborghs C.*, Roca Suarez AA., Van Renne N., Jühling F., Oudot MA, Virzì A., Bandiera S., Jamey C., Meszaros G., Brumaru D., Mukherjo A., Durand SC., Heydmann L., Verrier ER., El Saghire H., Hamdane N., Bartenschlager R., Fereshetian S., Ramberger E., Sinha R., Nabian M., Everaert C., Jovanovic M., Mertins P., Carr SA., Chayama K., Dali-Youcef N., Ricci R., Bardeesy NM., Fujiwara N., Gevaert O., Zeisel MB. Hoshida Y., Pochet N., Baumert TF. (2019) **Combined Analysis of Metabolomes, Proteomes, and Transcriptomes of Hepatitis C Virus–Infected Cells and Liver to Identify Pathways Associated With Disease Development.** *Gastroenterology* 157, 8

Poller W.C, Pieber M., Boehm-Sturm P., Ramberger E., Karampelas V., Möller K., Schleicher M., Wiekhorst F., Löwa N., Wahner S., Schnoor J., Taupitz M., Stangl K., Stangl V., Ludwig A. (2018) **Very small superparamagnetic iron oxide nanoparticles: Long-term fate and metabolic processing in atherosclerotic mice**. *Nanomedicine* 14, 8

Ramberger E., Dittmar G. (2017) **Tissue Specific Labelling in Proteomics**. *Proteomes,* 5, 17

Pamudurti N. R., Bartok O., Jens M., Ashwal-Fluss R., Stottmeister C., Ruhe L., Hanan M., Wyler E., Perez-Hernandez D., Ramberger E., Shenzis S., Samson M., Dittmar G., Landthaler M., Chekulaeva M., Rajewsky N., Kadener S. (2017) **Translation of CircRNAs**. *Mol. Cell* 66, 9–21.e7

Poller, W. C., Ramberger E., Boehm-Sturm P., Mueller S., Möller K., Löwa N., Wiekhorst F., Wagner S., Taupitz M., Schellenberger E., Baumann G., Stangl K., Stangl V., Ludwig A. (2016) **Uptake of citrate-coated iron oxide nanoparticles into atherosclerotic lesions in mice occurs via accelerated transcytosis through plaque endothelial cells**. *Nano Res.* 9, 3938

Bridges E., Sheldon H., Kleibeuker E., Ramberger E., Zois C., Barnard A., Harjes U., Li J., Masiero M., MacLaren R., Harris A., **RHOQ is induced by Dll4 and regulates angiogenesis by determining the intracellular route of the Notch intracellular domain.** *Submitted*

Acknowledgements

I want to thank my supervisors Prof. Dr. Achim Leutz and Prof. Dr. Gunnar Dittmar for giving me the opportunity to work on this interesting project and their support. I would also like to express my sincere gratitude towards Dr. Philipp Mertins for his support and feedback on my project.
I would like to thank the Berlin School of Integrative Oncology and the Max-Delbrück Center for Molecular Medicine for providing funding and a great environment for my PhD project.
I am grateful for all the help I got during my PhD, especially from Dr. Daniel Perez-Hernandez, Valeria Sapozhnikova, Dr. Elisabeth Kowenz-Leutz, Dr. Karin Zimmermann, Nathalie Nicot, Dr. Petr Nazarov and Arnaud Muller. I also want to thank all my other amazing colleagues in the proteomics platform at the MDC for great discussions, lending a hand whenever needed and fun times outside of the lab: Dr. Patrick Beaudette, Corinna Friedrich, Lorena Suarez-Artiles, Dr. Tamara Kanashova, Dr. Oliver Popp, Alina Dagane, Mohamad Haji, Sylvia Niquet, Dr. Marie-Luise Kirchner, Hans Werner and Merve Alp. Special thanks also go to all members of the Leutz lab at the MDC. I also want to thank Prof. Dr. Matthias Selbach and his group for fruitful discussions and input for my work. Last but not least I want to thank my family and friends for their never-ending support and encouragement. I am forever grateful to Tommaso for encouraging and helping me in every possible way.